DOCUMENT RESUME

ED 079 163 SE 016 574

AUTHOR	Dalton, LeRoy C., Ed.; Snyder, Henry D., Ed.
TITLE	Topics for Mathematics Clubs.
INSTITUTION	National Council of Teachers of Mathematics, Inc., Washington, D.C.
PUB DATE	73
NOTE	106p.
AVAILABLE FROM	NCTM, 1906 Association Drive, Reston, Virginia 22091 (No price quoted)
EDRS PRICE	MF-$0.65 HC Not Available from EDRS.
DESCRIPTORS	Algebra; *Enrichment Activities; Geometric Concepts; Instruction; *Mathematical Enrichment; Mathematics; Mathematics Education; Number Concepts; *Resource Materials; *Secondary School Mathematics; Topology

ABSTRACT

The ten chapters in this booklet cover topics not ordinarily discussed in the classroom: Fibonacci sequences, projective geometry, groups, infinity and transfinite numbers, Pascal's Triangle, topology, experiments with natural numbers, non-Euclidean geometries, Boolean algebras, and the imaginary and the infinite in geometry. Each chapter is written as a collection of related subtopics, and each includes a bibliography of references and further readings. (DT)

TOPICS FOR MATHEMATICS CLUBS

Topics for Mathematics Clubs

Edited by
LeRoy C. Dalton
Henry D. Snyder

NATIONAL COUNCIL OF TEACHERS OF MATHEMATICS
MU ALPHA THETA

Library of Congress Cataloging in Publication Data:

Main entry under title:

Topics for mathematics clubs.

Includes bibliographies.
1. Mathematics—Addresses, essays, lectures.
2. Mathematical recreations. I. Dalton, LeRoy C., 1928– ed. II. Snyder, Henry D., 1936– ed.
III. National Council of Teachers of Mathematics.
QA7.T63 510 73-6785

Printed in the United States of America

Contents

Preface

One of the main purposes of a mathematics club is to provide the opportunity and climate for students to study and present, before their peers, exciting topics in mathematics that are not ordinarily discussed in the classroom. It was with this in mind that the idea for this booklet was conceived.

Each of the chapters is intended to be a "turn on" treatment of a mathematical topic to interest students in that topic. Each is written as a collection of related subtopics; one result of this organization is that a committee of students can see the subtopic breakdown and each take a subtopic for presentation at a club meeting. The bibliographies are to suggest to the student where he can read in depth on his subtopic before presenting it to the club. Many sponsors have found that when three or more subtopics of a major topic are presented at one meeting, the students gain considerable feeling for what that area of mathematics is about.

The sponsor and/or student reading this booklet will note that many subtopics with their suggested references *individually* provide enough material for a club program and can be conveniently used in this manner. Still another way in which this booklet can be used is as a source for material for written projects for either a mathematics club or a mathematics course.

At a Mu Alpha Theta Governing Council Meeting held at the 1968 NCTM Meeting in Philadelphia, the idea of asking the NCTM to publish or jointly publish a book with materials for mathematics club programs was approved by the Mu Alpha Theta Council. A proposal was presented by Mu Alpha Theta to the NCTM Yearbook Planning Committee at its meeting in Chicago in October 1968. In November 1971 agreements were reached on all phases of this joint publication by the NCTM Publications Committee, the NCTM Board of Directors, and the Governing Council of Mu Alpha Theta.

The individuals in official capacities in the NCTM and Mu Alpha Theta

organizations most deserving of credit for their sincere efforts in making this joint publication possible are:

NCTM	*Mu Alpha Theta*
Arthur F. Coxford	Josephine P. Andree
M. Vere DeVault	Julius H. Hlavaty
Jack E. Forbes	Harold V. Huneke
James D. Gates	Robert L. Wilson
Charles R. Hucka	

Enough thanks cannot be given to the ten authors of the topics presented here. Although they all are very busy people, they responded enthusiastically when asked to write an article on a designated topic. Each did an excellent job with his writing assignment. The youth of America who read these articles should be most grateful for what they have done.

Also, our thanks are extended to Thomas Fitzpatrick and Henry Frandsen, who read certain of the manuscripts on request and gave some helpful suggestions.

Finally, our sincere thanks to Dorothy C. Hardy and the NCTM editorial staff for the excellent job done in preparing the manuscript for the printer.

LeRoy C. Dalton, *Editor*
Henry D. Snyder, *Associate Editor*

1

Fibonacci Sequences

Brother Alfred Brousseau

Intuitive Discovery of Fibonacci Relations

Fibonacci numbers got started with a problem dealing with the breeding of rabbits as found in the *Liber Abaci* (1202) of Leonardo Pisano (otherwise known as Fibonacci—one rendering of the meaning of this name being "son of good nature"). Assume, he says, that there is a pair of rabbits ready to breed and that each month they breed another pair of rabbits. However, rabbits that are bred in the first month do not breed during the second month. They begin to breed a pair a month only in the third month. Question: How many pairs of rabbits are there at the end of twelve months? Table 1.1 tells the story.

TABLE 1.1
THE BREEDING OF RABBIT PAIRS ACCORDING TO THE SCHEME SET UP BY FIBONACCI

Month	Rabbit Pairs Bred during Month	Pairs of Nonbreeding Rabbits	Pairs of Breeding Rabbits	Total No. of Pairs of Rabbits at End of Month
1	1	0	1	2
2	1	1	1	3
3	2	1	2	5
4	3	2	3	8
5	5	3	5	13
6	8	5	8	21
7	13	8	13	34
8	21	13	21	55
9	34	21	34	89
10	55	34	55	144
11	89	55	89	233
12	144	89	144	377

Now, what is so remarkable about this table? Looking at the numbers in the various columns it can be seen that each of the sequences of numbers is

part of the sequence: 0, 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, 233, 377, ... The distinctive property of this sequence is that each of the numbers (after the first two) is the sum of the two preceding numbers! Such is the origin of what is known as *the* Fibonacci sequence, which has since become famous for its many ramifications in the fields of mathematics, nature, and even technology.

To facilitate speaking about the sequence and its various terms, a simple type of notation is introduced: $F_0 = 0$, $F_1 = 1$, $F_2 = 1$, $F_3 = 2$, $F_4 = 3$, $F_5 = 5$, $F_6 = 8$, $F_7 = 13$, $F_8 = 21$, $F_9 = 34$, $F_{10} = 55$, and so on.

What characterizes a Fibonacci sequence is the recursion relation

$$T_{n+1} = T_n + T_{n-1},$$

which simply means that each term is the sum of the two preceding terms. One may start with any pair of integers, such as 2, 5, and build up a Fibonacci sequence: 2, 5, 7, 12, 19, 31, 50, 81, 131, ... However, there are two such sequences that have distinctive properties: *the* Fibonacci sequence: 1, 1, 2, 3, 5, 8, . . . mentioned above and the Lucas sequence: 1, 3, 4, 7, 11, 18, 29, 47, 76, 123, . . . , named after Edouard Lucas, a French mathematician of the latter half of the nineteenth century who did a great deal of work in connection with sequences of this and related types. For the Lucas sequence we use the notation $L_1 = 1$, $L_2 = 3$, $L_3 = 4$, $L_4 = 7$, $L_5 = 11$, $L_6 = 18$, $L_7 = 29$, $L_8 = 47$, ...

One of the fascinating aspects of Fibonacci sequences is this: There seems to be an unlimited opportunity to discover formulas and relations. For example, if each Fibonacci number is multiplied by its corresponding Lucas number we have the following:

n	F_nL_n
1	1
2	3
3	8
4	21
5	55
6	144

The products are Fibonacci numbers, being the Fibonacci numbers with even subscript. Intuitively, it seems that $F_nL_n = F_{2n}$.

Suppose we add the squares of two successive Fibonacci numbers. In carrying out such experiments, it is usually better to start higher up in the sequence than at the beginning, since often the results are masked by the presence of the two 1s at the beginning of the sequence. Let us try a few pairs and see what result we obtain.

$$F_3^2 + F_4^2 = 2^2 + 3^2 = 13, \quad F_4^2 + F_5^2 = 3^2 + 5^2 = 34, \quad F_5^2 + F_6^2 = 5^2 + 8^2 = 89, \ldots.$$

Again the result is a Fibonacci number, and actually, the Fibonacci number whose subscript is the sum of the subscripts of the two squared numbers. That is, we have a general relation

$$F_n^2 + F_{n+1}^2 = F_{2n+1}.$$

It is a good idea, once we have a Fibonacci relation, to see what happens

when we perform a corresponding operation with the Lucas sequence. Try this in the present case.

Is there some way of tying in the Lucas sequence with the Fibonacci sequence? Examine what happens when alternate terms of the Fibonacci sequence are added together. For example, the first and third add up to 3, the second and fourth add up to 4, the third and fifth to 7, the fourth and sixth to 11, and so on. These are all Lucas numbers, the Lucas number lying between the two Fibonacci numbers that are being summed. Hence the general relation is

$$L_n = F_{n+1} + F_{n-1}.$$

Try the corresponding relation adding alternate Lucas numbers.

An interesting situation arises when we have a series of Fibonacci relations that lead to an overall general formula. For example, try the experiment of taking any Fibonacci number squaring it, and comparing this square to the product of the Fibonacci numbers on either side of it. The difference is always one. However, in some cases it is plus one and in others a minus one. The entire result can be summarized in the following formula:

$$F_n^2 - F_{n-1}F_{n+1} = (-1)^{n-1}.$$

Now compare the square of any Fibonacci number to the product of the numbers two away on either side. Again the difference is one, but in the opposite sense. so that

$$F_n^2 - F_{n+2}F_{n-2} = (-1)^n.$$

The road from here on is clear. Compare the square with the product of the Fibonacci numbers three away on either side. The difference is always four, and the relation obtained is:

$$F_n^2 - F_{n+3}F_{n-3} = 4(-1)^{n-1}.$$

Pursue this sequence. Ultimately one should arrive at a formula for

$$F_n^2 - F_{n+k}F_{n-k}.$$

Try the same sequence of formulas using Lucas numbers.

Another type of relation that may be investigated is various sums: for example, the sum of the terms of the Fibonacci sequence. Again, avoiding for the moment the early cases, one finds that the sum of the first five terms is 12; the sum of the first six terms is 20; the sum of the first seven terms is 33; the sum of the first eight terms is 54; and so on. Invariably the sum is one less than a Fibonacci number, the number that is two steps beyond the last number summed. Hence one surmises that

$$\sum_{k=1}^{n} F_k = F_{n+2} - 1.$$

What is the sum of the first n Lucas numbers?

Other sums that may be investigated in both sequences are the sum of the odd-subscript terms, the sum of the even-subscript terms, the sum of the squares, the sum of terms where the subscripts differ by four, and so on.

For more leads and hints about intuitive discovery see *Introduction to*

Fibonacci Discovery by the author and his mimeographed publication entitled "Fibonacci Formulas Suitable for Intuitive Discovery." For historical background, see Gies and Gies.

Fibonacci Sequences and the Golden Section Ratio

The "golden section" is a famous piece of geometry that is found in Euclid's *Elements*. The problem is to take a line segment AB and to place a point C so that $AB:AC = AC:BC$, as seen in figure 1.1. In other words, the larger section of the line segment AB is the mean proportional between the whole line segment and the smaller section. What is the value of this ratio? Let $AB/AC = x$. It is convenient to take $AB = x$ and $AC = 1$, since this gives the required ratio. Then $BC = x - 1$. Hence

$$x:1 = 1:x - 1,$$

from which $x^2 - x - 1 = 0$. Solving, there are two roots:

$$r = \frac{1+\sqrt{5}}{2} \quad \text{and} \quad s = \frac{1-\sqrt{5}}{2}$$

The first is approximately equal to 1.6180339887 and the second to -0.6180339887. The positive value is the ratio AB/AC, the golden section ratio. (Sometimes its reciprocal, 0.6180339887, is taken as such.)

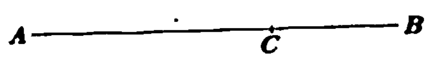

Fig. 1.1

Now one of the fascinating things about Fibonacci numbers is the fact that they are strongly related to this golden section ratio. We arrive at this relation when we try to answer the question: Is there an explicit formula for the terms of the Fibonacci sequence? The answer is yes, and the solution brings out a beautiful piece of algebra.

We start with the recursion relation for a Fibonacci sequence

$$T_{n+1} = T_n + T_{n-1} \quad \text{or} \quad T_{n+1} - T_n - T_{n-1} = 0\,.$$

We form a corresponding quadratic equation with the same coefficients:

$$x^2 - x - 1 = 0,$$

which is evidently the same as the equation for obtaining the golden section ratio with roots r and s as given above. Now, since r is a root of this equation,

$$r^2 - r - 1 = 0 \quad \text{or} \quad r^2 = r + 1\,.$$

Multiplying through by r^{n-1}, we have

$$r^{n+1} = r^n + r^{n-1}.$$

Note that the powers of the roots have the same recursion relation as the terms of the Fibonacci sequence. This gives rise to the idea of expressing the terms of a Fibonacci sequence in the form

$$T_n = Ar^n + Bs^n,$$

where A and B are suitable constants. If this form holds up to T_n, then since $T_{n+1} = T_n + T_{n-1}$,

$$\begin{aligned} T_{n+1} &= Ar^n + Bs^n + Ar^{n-1} + Bs^{n-1} \\ &= A(r^n + r^{n-1}) + B(s^n + s^{n-1}) \\ &= Ar^{n+1} + Bs^{n+1}. \end{aligned}$$

Hence if this form can be established for the first two terms of the Fibonacci sequence, for example, it will continue for all terms. To do this we set

$$1 = Ar + Bs \quad \text{and} \quad 1 = Ar^2 + Bs^2,$$

the solution for A and B being $A = 1/\sqrt{5}$ and $B = -1/\sqrt{5}$. Thus an explicit expression for the terms of the Fibonacci sequence is:

$$F_n = \frac{r^n - s^n}{\sqrt{5}}.$$

Proceeding similarly for the Lucas sequence, one finds

$$L_n = r^n + s^n.$$

These formulas, known as the Binet formulas for the Fibonacci and Lucas numbers, are not particularly practical for calculating such numbers, but they provide powerful tools for developing and proving Fibonacci relations.

For example,

$$F_{2n} = \frac{r^{2n} - s^{2n}}{\sqrt{5}} = \frac{(r^n - s^n)(r^n + s^n)}{\sqrt{5}} = F_nL_n,$$

and we have quickly proved one of the relations we found by intuition.

Again

$$\begin{aligned} F_n^2 - F_{n-1}F_{n+1} &= \frac{(r^n - s^n)^2}{5} - \frac{(r^{n-1} - s^{n-1})(r^{n+1} - s^{n+1})}{5} \\ &= \frac{(r^{2n} - 2r^ns^n + s^{2n} - r^{2n} + r^{n-1}s^{n+1} + s^{n-1}r^{n+1} - s^{2n})}{5} \\ &= \frac{r^{n-1}s^{n-1}(r^2 - 2rs + s^2)}{5} \end{aligned}$$

Now $r^2 + s^2 = L_2 = 3$ and $-2rs = -2(-1) = 2$. Hence

$$F_n^2 - F_{n-1}F_{n+1} = (-1)^{n-1}.$$

A fascinating aspect of the Fibonacci sequence is the fact that if we take the ratio of consecutive terms, the value gets closer and closer to the golden section ratio as we go out in the sequence.

n	$\frac{F_{n+1}}{F_n}$	n	$\frac{F_{n+1}}{F_n}$
1	1	6	1.625
2	2	7	1.6153846153
3	1.5	8	1.6190476190
4	1.666666666 . . .	9	1.6176470588
5	1.6	10	1.6181818181

For $n = 20$, the ratio is 1.6180339985 .

But what is true of *the* Fibonacci sequence is true of any Fibonacci sequence. For example, for the sequence 2, 5, 7, 12, . . . , the ratio of the twenty-first term to the twentieth is

$$\frac{42{,}187}{26{,}073} = 1.6180339815.$$

Why is this? Let us examine this ratio for the Fibonacci sequence.

$$\frac{F_{n+1}}{F_n} = \frac{\dfrac{r^{n+1} - s^{n+1}}{\sqrt{5}}}{\dfrac{r^n - s^n}{\sqrt{5}}}$$

$$= \frac{r^{n+1} - s^{n+1}}{r^n - s^n}.$$

Dividing the terms in numerator and denominator by r^n, one has

$$\frac{F_{n+1}}{F_n} = \frac{r - s\left(\frac{s}{r}\right)^n}{1 - \left(\frac{s}{r}\right)^n}.$$

But s/r has a value less than one, and hence when n gets larger this fraction to the nth power gets closer and closer to zero so that the limiting ratio is r.

Fibonacci Numbers in the World

People are used to finding mathematics embodied in their culture because man has put it there. But when they find something so elegant as the Fibonacci sequence or the golden section ratio in nature, they begin to wonder whether somehow mathematics is not an integral part of the plan of the universe.

Perhaps the most striking and common instances are found in plants. It has long been recognized by botanists that the leaf arrangement (phyllotaxis) of many plants follows a Fibonacci pattern. This pertains to plants where the leaves spiral up the stem. Take a leaf and find the next leaf up the stem which is vertically above it. Counting the original leaf as the zero leaf, find how many steps there are to the next leaf vertically above the zero leaf. Very often this is a Fibonacci number, such as 3, 5, 8, 13, ... Now determine how many times it was necessary to go around the stem in order to arrive at the leaf. For five steps, this is two, for eight it is three, and so on, in Fibonacci arrangements. Such arrangements are expressed by the following ratios: 1/2, 1/3, 2/5, 3/8, 5/13, 8/21, and so on. The meaning of this notation is twofold: 3/8 means that it takes three revolutions and eight steps to get to the leaf vertically above the zero leaf. Or another way of thinking about this would be: To go from one leaf to the next leaf in the sequence takes on the average 3/8 of a revolution around the stem, since eight steps give three revolutions. Note that these numbers are alternate Fibonacci numbers. The 1/2 arrangement is the simple arrangement of what is known as alternate leaves and fits into the general pattern.

If we make a table of angles corresponding to the various cases we have table 1.2.

TABLE 12

Phyllotaxis Arrangement	Angle
1/2	180°
1/3	120°
2/5	144°
3/8	135°
5/13	138.5°
8/21	137.1°
13/34	137.6°

We seem to be approaching a limiting angle. This is because the ratio of alternate terms of the Fibonacci sequence as given is approaching $1/r^2$, and the limiting angle corresponding to this is 137.49°.

Modified leaf arrangements are found in the bracts of pine cones and the seed arrangements of sunflowers. A somewhat different way of counting brings out the Fibonacci numbers in these cases. For pine cones, there are usually two obvious spirals, one steeper and the other more gradual. One way of counting is to determine how many steep spirals there are and how many gradual spirals. These are found to be consecutive Fibonacci numbers, such as 8 and 5. A second way of counting is to start from any one bract and follow the two spirals going through it to their next intersection. Counting the number of steps along each spiral gives the same two consecutive Fibonacci numbers.

The author has counted thousands of pine cones of various species, as well as cones of Douglas fir, redwood, and spruces of various kinds. Up to approximately 99 percent, the cones of one species have a common Fibonacci pattern. Deviations give either Lucas counts or double or triple Fibonacci and Lucas counts. For example, a count of six and ten would be double three and five. Thus, almost without exception there is some Fibonacci or Lucas count on the cones of conifers.

Cacti of certain species (*Opuntia*, for example) show spirals and Fibonacci or Lucas counts. Usually, however, there seems to be more variety than among pine cones. Furthermore, the internal wooden structure that remains after the cactus dies is a weblike pattern of spirals that gives Fibonacci and Lucas counts inasmuch as the spines come through the holes in the webbing.

Interesting patterns have also been found in other desert plants such as the ocotillo and the Joshua tree.

The pineapple is a prime example of Fibonacci patterns, there being as many as four spirals through one bract, each pair of spirals interacting to give Fibonacci numbers. Multiple spirals can also be found on pine cones.

Perhaps the prize example of Fibonacci numbers in nature is the sunflower. The seeds on the head are arranged in spirals, usually a steep and a not-so-steep spiral being evident. Counting the number of steep and the number of gradual spirals gives such numbers as 89 and 55, 55 and 34, 76 and 47. One may also count a third set of spirals going rather directly toward the center. In the case of 89 and 55, this set would give a number

144; for 76 and 47, the number of spirals toward the center would be 123 (tenth Lucas number).

Another example of Fibonacci numbers in nature is found in the ancestry of the male bee. Male bees hatch from an unfertilized egg; female bees, from a fertilized egg. If we trace back the number of ancestors of a male bee (from the bottom to the top in figure 1.2), we have a pattern that produces Fibonacci numbers.

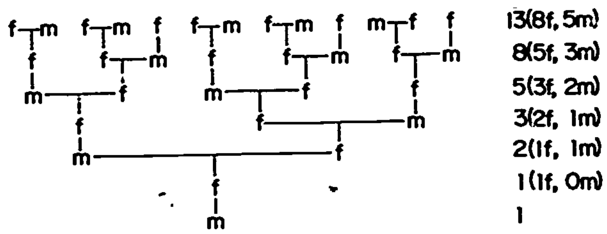

Fig. 1.2

There are some indications that Fibonacci numbers and the golden section ratio are related to the structure of atoms and the spacing of the planets in the solar system.

When we consider the world that man has created, there are many indications going back to ancient times of the conscious or unconscious use of Fibonacci numbers and the golden section ratio. For example, Donald A. Preziosi finds evidence of Fibonacci ratios and the golden section ratio in Minoan architecture. Richard E. M. Moore, after making extensive studies of mosaics, came to the conclusion that some basic Fibonacci units were being employed by the people making them. The golden section ratio is evident in Greek architecture, and many painters employed it in the construction of their designs.

Fibonacci sequences have also been used in the creation of music, a notable example being Krenek's "Fibonacci Mobile."

In modern times there are those who believe (Ellis Wave Theory) that Fibonacci numbers have something to do with the fluctuations of the stock market, though they hasten to add that using this information would not enable any one to become rich.

Finally, these numbers have been found useful in computer science and in search techniques.

REFERENCES

Note.—References written by the author of this chapter or pertaining to the Fibonacci Association may be obtained by writing to Brother Alfred Brousseau, Saint Mary's College, Moraga, California 94575. No charge is made for mimeographed materials.

INTUITIVE DISCOVERY OF FIBONACCI RELATIONS

Brousseau, Brother Alfred. *An Introduction to Fibonacci Discovery*. St. Mary's College, Calif.: Fibonacci Assn., 1965.

———. "Fibonacci Formulas Suitable for Intuitive Discovery." Mimeographed.

Gies, Joseph, and Frances Gies. *Leonard of Pisa*. New York: Thomas Y. Crowell Co., 1970.

Hoggatt, Verner E., Jr. *Fibonacci and Lucas Numbers*. Boston: Houghton Mifflin Co., 1969.

FIBONACCI SEQUENCES AND THE GOLDEN SECTION RATIO

Brousseau, Brother Alfred. *An Introduction to Fibonacci Discovery*. St. Mary's College, Calif.: Fibonacci Assn.. 1965.

———. *Linear Recursion and Fibonacci Sequences*. St. Mary's College, Calif.: Fibonacci Assn., 1971.

Hoggatt, Verner E., Jr. *Fibonacci and Lucas Numbers*. Boston: Houghton Mifflin Co.. 1969.

FIBONACCI NUMBERS IN THE WORLD

Basin, S. L. "The Fibonacci Sequence As It Appears in Nature." *Fibonacci Quarterly*, February 1963, pp. 53–56.

Beard, Robert S. "Design of the Great Pyramid of Gizeh." *Fibonacci Quarterly*, February 1968, pp. 85–87.

Brousseau, Brother Alfred. "Fibonacci Numbers in Nature." Mimeographed.

———. "Fibonacci Statistics in Conifers." *Fibonacci Quarterly*, December 1969, pp. 525–32. Also mimeographed.

———. "On the Trail of the California Pine." *Fibonacci Quarterly*, February 1968, pp. 69–76. Also mimeographed.

"The Fibonacci Numbers." *Time*, 4 April 1969, pp. 48 and 50.

Gardner, Martin. "A Discussion of Helical Structures, from Corkscrews to DNA Molecules." *Scientific American*, June 1963, p. 152.

———. "The Multiple Fascinations of the Fibonacci Sequence." *Scientific American*, March 1969, pp. 116–20.

Karchmar, E. J. "Phyllotaxis." *Fibonacci Quarterly*, February 1965, pp. 64–66.

Lowman, Edward A. "An Example of Fibonacci Numbers Used to Generate Rhythmic Values in Modern Music." *Fibonacci Quarterly*, October 1971, pp. 423–26.

———. "Some Striking Proportions in the Music of Bela Bartok." *Fibonacci Quarterly*, December 1971, pp. 527-28, 536.

McNabb, Sister Mary de Sales. "Phyllotaxis." *Fibonacci Quarterly*, December 1963, pp. 57–60.

Read, B. A. "Fibonacci Series in the Solar System." *Fibonacci Quarterly*, October 1970, pp. 428–38.

Moore, Richard E. M. "Mosaic Units: Pattern Sizes in Ancient Mosaics." *Fibonacci Quarterly*, April 1970, pp. 281–310.

Norden, H. "Proportions in Music." *Fibonacci Quarterly*, October 1964, pp. 219–22.

Preziosi, Donald A. "Harmonic Design in Minoan Architecture." *Fibonacci Quarterly*, December 1968, pp. 370–84.

Wlodarski, J. "The 'Golden Ratio' and the Fibonacci Numbers in the World of Atoms." *Fibonacci Quarterly*, December 1963, pp. 61–64.

2

Projective Geometry

Donald J. Dessart

What Is Projective Geometry?

In the nineteenth century Arthur Cayley (1821–1895), a prominent mathematician, said, "Projective geometry is all geometry." More recently a twentieth-century mathematician, Morris Kline, wrote, "In the house of mathematics there are many mansions and of these the most elegant is projective geometry" (1956, p. 622). Many scholars of both centuries agreed that projective geometry is one of the most fascinating branches of mathematics to study because of its beautiful and surprising properties.

The roots of projective geometry can be traced to artists of the Renaissance. Such famous painters as Leonardo da Vinci and Raphael and such architects, as Brunelleschi were apparently well aware of notions of projective geometry, as evidenced by their creations. The basic problem of an artist is one of capturing three-dimensional reality on a two-dimensional canvas. To illustrate this problem, imagine that you are an artist peering through a window into a massive hall with circular pictures on the walls and with a floor made up of large, brightly colored square tiles. Imagine that you wish to recreate this scene by sketching it on the surface of the windowpane through which you are looking. In accomplishing this, you might think of straight lines projecting from various points of the scene through the window and to your eye; and wherever a line of light passes through the windowpane, you would mark a dot. If you were able to do this for many of the lines, the mass of dots could provide a crude picture of the scene. (One might observe that this is essentially what a camera does when it focuses light from a scene through the lens of the camera onto the film.)

As the scene emerges on the windowpane, you would observe that the squares of the tiled floor would not be represented by squares on the window-

pane but would be distorted into various kinds of quadrilaterals. The right angles of the squares would not usually appear as right angles but as obtuse or acute angles. Circular objects on the walls would usually appear as elliptical shapes, and other kinds of distortions would occur.

You might also observe that lines that are straight lines in the scene would all appear as straight lines on the windowpane; and that points "on" certain lines of the scene would remain "on" the same lines of the sketch. The number of sides of the figures on the tiled floor would equal the number of sides of the corresponding figures on the windowpane. In other words, you could conclude that certain properties such as the size of the angles or the lengths of the sides of figures would usually be changed by the sketching, whereas other properties, such as the number of sides of figures, would remain unchanged. You would have observed what is the essence of the study of projective geometry because this geometry may be described as the study of those properties of figures that persist or remain unchanged by projections.

Those mathematicians who studied and developed projective geometry had to introduce ideas which at first may seem quite strange. For example, Gerard Desargues (1593–1662), an early researcher, found that in projective geometry it was desirable to consider parallel lines as having a common point at an infinite distance. This point in which the parallel lines "meet" is called an "ideal point" of projective geometry; similarly, parallel planes can be considered as meeting at a distant line, which is called an "ideal line." So, in the plane of projective geometry *any* two lines have a point in common; and in three-space of projective geometry *any* two planes have a line in common! In a further study of this subject, you will find many other strange and fascinating ideas.

Principle of Duality

We all like to be efficient and enjoy avoiding unnecessary work. Mathematicians are certainly no different from others in this respect. In fact, when a mathematician discovers a method or a principle that saves him a great deal of time and energy, he frequently refers to such a principle as being "beautiful," in the sense that it is very efficient. Such a beautiful principle in projective geometry is the *principle of duality in the plane,* which we will first state and then illustrate by examples. You may wish to make up other examples; and the members of your mathematics club should be able to think of some, too.

This is the principle of duality in the plane: *If any statement involving points and lines of a projective plane, usually stated in the "on" language, is true, then another true statement is obtained by interchanging the words "point" and "line" of the first statement.*

For example, we know that in Euclidean geometry (the geometry studied in most high schools) any two points determine one and only one line, or stated in "on" language, that any two points lie on one and only one line.

Suppose we "dualize" this statement by employing the principle of duality in the plane. The dual of "Any two *points* lie on one and only one *line*" is the statement, "Any two *lines* lie on one and only one *point*." Both of these statements are true in projective geometry, whereas the first is true but the second is not true in Euclidean geometry.

Two other dual concepts are provided by collinear points and concurrent lines. Any set of points on the same line, such as A, B, C, D in figure 2.1 are said to be *collinear* because they all lie on the line l; in fact, all of the points of the line are collinear. Similarly, lines such as a, b, c, d in figure 2.2 are said to be *concurrent* because they all lie on the point E, and all of the lines that are on the point E are concurrent. From this we can see that *points* being collinear and *lines* being concurrent are dual notions in projective geometry.

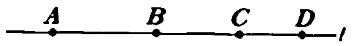

Fig. 2.1. Collinear points

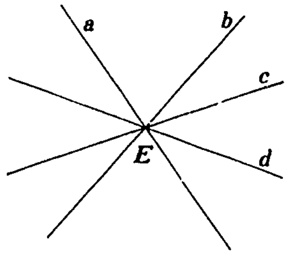

Fig. 2.2. Concurrent lines

As another example, consider a figure that consists of four *points:* L, M, N, O, as shown in figure 2.3, such that no three of the four *points* are collinear. The dual of this figure consists of four *lines:* l, m, n, o, such that no three of the four *lines* are concurrent. (See fig. 2.4.)

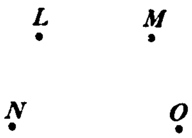

Fig. 2.3

Suppose that we defined a "triangle" as a figure consisting of the union of three noncollinear points and the three lines that are determined by these points, taken a pair at a time. (Note that this is not the usual definition of a triangle because it uses lines rather than line segments.)

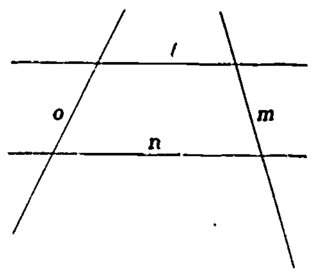

Fig. 2.4

Carefully dualize this definition by interchanging "point" and "line" and "concurrent" and "collinear," wherever these words appear. Surprisingly, you will find that the dual statement describes the same configuration and is therefore another description of a "triangle." This figure is an example of a *self-dual* figure of projective geometry.

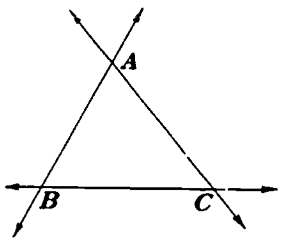

Fig. 2.5. A triangle

We can see that the principle of duality in the plane provides for efficiency because in proving one theorem, we are also proving its dual, which may be another theorem. It is a real bargain, since we are getting two things for the price of one!

Desargues's Theorem

When one looks at a structure in the distance, such as a tall building, two images of the building, one for each eye, are sent as sensations to the brain. When the brain compares these two sensations, it provides a perception of the relative depth of the various parts of the structure. On the other hand, if one were to look at the building with only one eye, then essentially a flat or two-dimensional picture is sent to the brain. Apparently, in this case, the brain is able to compensate to a certain extent in order to provide the viewer with some impression of depth; but not as well as it can when it receives two sensations.

A very remarkable example of this compensation by the brain is seen in the story of Wiley Post, who was a pioneer in high-altitude flight in the United States. When Post was twenty-five years old, he lost his left eye in

an oil-field accident. Using the insurance money he received for the accident, he purchased his first airplane; and in spite of being seriously handicapped, he became one of America's most famous aviators. His exploits are described more fully in the Smithsonian Institution in Washington, D.C., which displays a cumbersome tank suit that he wore while flying at high altitudes.

Suppose that someone with only one eye, such as Wiley Post, or a person with normal vision who has covered one eye, were to view triangle ABC from the point O and that flat, translucent screens were then placed between the viewer and triangle ABC so that triangles $A'B'C'$ and $A''B''C''$ would be formed on the screens by straight lines of light such as those through O and points A, B, C, respectively. (See fig. 2.6.) We say that the triangles $A'B'C'$ and $A''B''C''$ are *perspective from the point*, O, and, in general, two triangles are perspective from a point if their vertices can be placed into one-to-one correspondence so that lines joining corresponding vertices pass through that point, called the center of perspectivity.

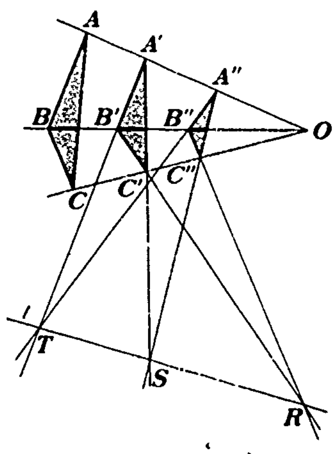

Fig. 2.6

Desargues made the astute observation that if one extended the sides of the triangles $A'B'C'$ and $A''B''C''$ so that $\overleftrightarrow{A'B'}$ and $\overleftrightarrow{A''B''}$ would intersect in T, $\overleftrightarrow{A'C'}$ and $\overleftrightarrow{A''C''}$ would intersect in S, and $\overleftrightarrow{B'C'}$ and $\overleftrightarrow{B''C''}$ would intersect in R, that the intersection points R, S, T would be collinear! The line l, containing R, S, T is called the axis of perspectivity, and triangles $A'B'C'$ and $A''B''C''$ are said to be perspective from the line, l. In general, two triangles are perspective from a line if the sides of the triangles can be placed into one-to-one correspondence so that points of intersection of corresponding sides lie on a straight line.

The theorem of Desargues can be stated as follows: *If two triangles are perspective from a point, then the two triangles are perspective from a line.*

In the case we just examined, the two triangles $A'B'C'$ and $A''B''C''$ were

in different planes (the planes of the screens), but this theorem is also true for two triangles in the same projective plane.[1] You may wish to experiment with this case by drawing two triangles on a sheet of paper which are perspective from a point.

Furthermore, if the "two for the price of one" principle of plane duality is applied to Desargues's theorem, one will obtain another theorem, which states, *"If two triangles are perspective from a line, then the two triangles are perspective from a point."* This is the converse of Desargues's theorem and is true for triangles in the projective plane as well as in three-space.

If members of your mathematics club become interested in this theorem, they may wish to attempt to prove it. As a hint, it is much easier to prove the theorem for two triangles in three-space than for two triangles in the projective plane. Some of the references listed at the end of this article supply such proofs.

Pascal's Theorem

Blaise Pascal (1623–1662), when only sixteen years old, proved a theorem known today as Pascal's theorem, that amazed older and more experienced mathematicians of his time. As a youngster, he also "discovered," by folding a triangle cut from paper, that the sum of the angles of a triangle is a straight angle (members of your club may wish to experiment with this idea by folding triangles cut from paper). In addition to these remarkable achievements, Pascal demonstrated his genius in many other fields. He designed and built an adding machine, solved problems in hydrostatics, and wrote scholarly papers in arithmetic, algebra, probability, the theory of numbers, and theology.

Before considering Pascal's theorem, we need to describe what we will mean by a plane hexagon. Consider a figure consisting of six coplanar points, A_1, A_2, A_3, A_4, A_5, A_6, no three of which are collinear, and taken in numerical order of the subscripts to form six line segments joining pairs of successive points. The union of the six points, called vertices, and the six line segments, called sides, is a *plane hexagon*. Plane hexagons may look like any of the shapes in figure 2.7.

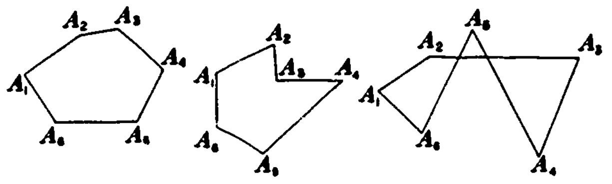

Fig. 2.7. Plane hexagons

1. Perhaps it should be pointed out that Desargues's theorem does not hold true in all projective planes. For a discussion of non-Desarguesian geometries, see Dorwart, pp. 121–30; Eves, pp. 362–65; and Pedoe, pp. 29–32.

Suppose that a plane hexagon $A_1, A_2, A_3, A_4, A_5, A_6$ is inscribed in a circle as shown in figure 2.8, and that a pair of opposite sides ($\overline{A_1A_2}$ and $\overline{A_4A_5}$) are extended so that they intersect in a point, P, another pair of opposite sides ($\overline{A_2A_3}$ and $\overline{A_5A_6}$) intersect in a point, R, and the third pair of opposite sides ($\overline{A_3A_4}$ and $\overline{A_6A_1}$) intersect in a point, Q, then Pascal observed that the points Q, P, R would lie on a straight line. Pascal was able to prove that if any plane hexagon is inscribed in a circle, then the opposite sides of the plane hexagon (extended, if necessary), intersect in three points that are collinear.

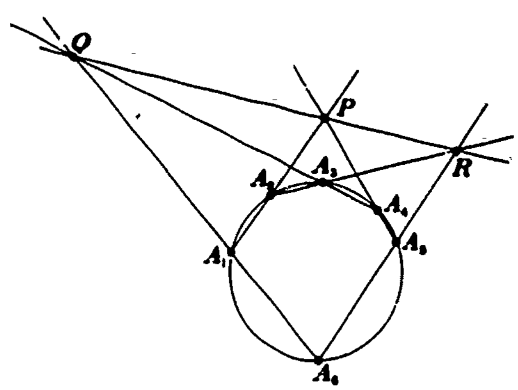

Fig. 2.8

There are many other fascinating consequences of this theorem. Suppose we take any six points on a circle. It can be shown that 60 different plane hexagons could be formed. To each of these 60 hexagons there corresponds a line, given by Pascal's theorem, called a Pascal line. It can be demonstrated that these 60 lines pass, three by three, through a total of 20 points, called Steiner points; and these 20 points lie, four by four, on 15 lines, named Plücker lines. There are many other extensions of Pascal's theorem which make it live up to its earlier name of the "Mystic Hexagram Theorem."

Cross Ratio

Imagine, again, that you are an artist (see fig. 2.9) looking through a windowpane, m, from the point O, at the line, l, which contains the points R, S, T, U, as shown; and R', S', T', U' are images of the points R, S, T, U (respectively) formed on the windowpane by lines of light which pass through O. One can observe that the lengths of the line segments, $\overline{RS}$ and $\overline{R'S'}$, do not appear to be equal. If we designate these lengths by RS and $R'S'$, it seems reasonable to conclude that $RS \neq R'S'$. Consequently, it seems clear that the lengths of other line segments of l would not be equal to the corresponding line segments of m under a projection, such as the one shown.

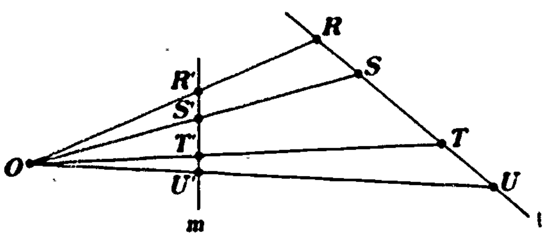

Fig. 2.9

One might wonder if the ratios of lengths of corresponding segments under a projection are equal. For example, one might ask, "Is RT/TS equal to $R'T'/T'S'$?" The answer to this question, while not obvious, is, "No, because there are projections for which $RT/TS \neq R'T'/T'S'$." At this point, many people would be inclined to give up searching for relationships between corresponding segments; but the early mathematicians did not give up and found that there is a ratio, or more correctly, a ratio of ratios that is not altered by projection. This ratio, named earlier in history a "double ratio," is known today as a *cross ratio*. If one will allow the length, RS, to be regarded as positive, and the length, SR, to be negative, that is, if we are willing to introduce the notion of directed line segments, then the cross ratio may be defined as follows: *If R, S, T, U are four distinct points on a line, the ratio of ratios, $(RT/TS)/(RU/US)$, designated by the symbol (RS, TU) is the cross ratio of R, S, T, and U, in that order.*

The ancient Greek mathematicians discovered that in such a projection as the one described, (RS, TU) would equal $(R'S', T'U')$; that is,

$$\frac{RT/TS}{RU/US} = \frac{R'T'/T'S'}{R'U'/U'S'}.$$

They observed an extremely important property of projective geometry; namely, that cross ratio remains unaltered, or is invariant, under projection. Because the cross ratio has this property, it plays a very special role in projective geometry; and if one were to study this subject in depth, he would become extremely familiar with the idea of a cross ratio.

We can study a few significant properties of cross ratio without delving deeply into the subject. We have seen that the definition of cross ratio relies upon the particular order in which the points R, S, T, U, are considered. If we select the points in another order, such as, S, R, T, U, and then apply the definition of cross ratio, we obtain $(SR, TU) = (ST/TR)/(SU/UR)$. As we shall see later, $(SR, TU) \neq (RS, TU)$. Suppose that we select the points in the order S, R, U, T. Applying the definition of cross ratio, we find that $(SR, UT) = (SU/UR)/(ST/TR)$. In this case, it can be shown that $(SR, UT) = (RS, TU)$. Let's see how!

Since RT and TR are of the same absolute length but have opposite signs, and similarly for TS and ST, then

$$\frac{RT}{TS}\cdot\frac{ST}{TR} = 1,$$

and, similarly,

$$\frac{RU}{US}\cdot\frac{SU}{UR} = 1.$$

Thus, we can conclude that

$$\frac{RT}{TS}\cdot\frac{ST}{TR} = \frac{RU}{US}\cdot\frac{SU}{UR};$$

and, furthermore, that

$$\frac{RT/TS}{RU/US} = \frac{SU/UR}{ST/TR};$$

or

$$(RS, TU) = (SR, UT).$$

It can be easily seen that there are 4! or 24 permutations of the points R, S, T, U, taken four at a time. It will prove helpful at this time if you will enumerate these 24 permutations. Each of them gives rise to a particular value of the cross ratio, and one might wonder if the patterns of selections of points determine these values. This is, indeed, the case! If we let $(RS, TU) = k$, then each of the other 23 cross ratios is equal either to $k, 1/k, 1-k, 1/(1-k), (k-1)/k$, or $k/(k-1)$.

In a further study of these patterns, it can be seen that there are three basic rules involving changes in the order of the points. Each change affects the value of the cross ratio. If we let $(RS, TU) = k$ in this discussion, these rules, which are easily justified (see Eves, pp. 73–74), can be expressed as follows:

R(1): An interchange of any two points combined with an interchange of the remaining two points does not alter the value of the cross ratio; for example, if we interchange R and S and also interchange T and U, we obtain $(RS, TU) = (SR, UT) = k$.

R(2): An interchange of only the first pair of points changes the value of the cross ratio from k to $1/k$, for example, $(SR, TU) = 1/k$.

R(3): An interchange of only the middle pair of points changes the value of the cross ratio from k to $1-k$; for example, $(RT, SU) = 1-k$.

Applying these rules successively on cross ratios, we may see how each of the six values $k, 1/k, (k-1)/k, k/(k-1), 1/(1-k)$, and $1-k$ may be obtained.

1. Start with $(RS, TU) = k$.
2. Applying R(2) to (RS, TU), we obtain $(SR, TU) = \frac{1}{k}$.
3. Applying R(3) to (SR, TU), we obtain $(ST, RU) = 1 - \frac{1}{k} = \frac{k-1}{k}$.
4. Applying R(2) to (ST, RU), we obtain $(TS, RU) = \frac{k}{k-1}$.

5. Applying R(3) to (TS, RU), we obtain $(TR, SU) = 1 - \frac{k}{k-1} = \frac{1}{1-k}$.

6. Applying R(2) to (TR, SU), we obtain $(RT, SU) = 1 - k$.

Furthermore, each of the 24 possible cross ratios described above will fall into one of the six categories of equal cross ratios given below; in the first category each cross ratio is equal to k, in the second to $1/k$, and so forth—in the last category to $k/(k-1)$.

a. $(RS, TU) = (SR, UT) = (TU, RS) = (UT, SR) = k$.

b. $(SR, TU) = (RS, UT) = (UT, RS) = (TU, SR) = \frac{1}{k}$.

c. $(RT, SU) = (SU, RT) = (TR, US) = (US, TR) = 1 - k$.

d. $(TR, SU) = (US, RT) = (RT, US) = (SU, TR) = \frac{1}{1-k}$.

e. $(ST, RU) = (RU, ST) = (UR, TS) = (TS, UR) = \frac{k-1}{k}$.

f. $(TS, RU) = (UR, ST) = (RU, TS) = (ST, UR) = \frac{k}{k-1}$.

Harmonic Set of Points

The complete quadrilateral provides an interesting application of cross ratio. It is a figure consisting of the union of four lines in a plane, no three of which are concurrent, and the six points in which pairs of these lines intersect. The four lines are named sides of the quadrilateral and the six points are called its vertices. Two vertices, such as A and C in figure 2.10, which do not lie on the same line are called opposite vertices. The lines joining pairs of opposite vertices, sh[illegible] by the dotted lines, are the diagonals of the complete quadrilateral. (Your club members may wish to dualize this [illegible]inition; the resulting statement is a definition of a figure called a complete quadrangle).

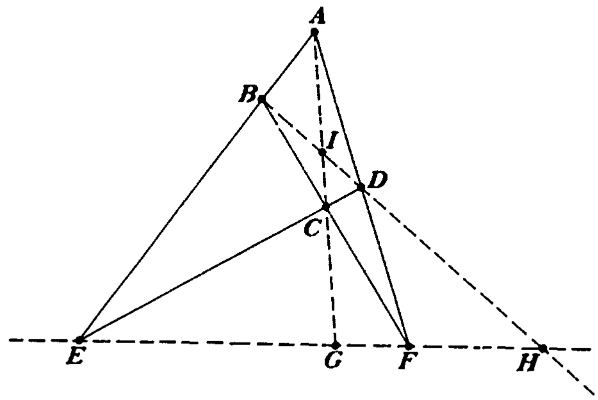

Fig. 2.10

If any diagonal of the complete quadrilateral, such as $\overleftrightarrow{EF}$, is taken and the points of intersection with the other diagonals, G and H, are determined, the set of points E, G, F, H is called a ***harmonic set of points***. The cross ratio of any harmonic set of points is equal to -1; so, in this case $(EF, GH) = -1$.

We can demonstrate rather easily that $(EF, GH) = -1$. First, we need to recall from our earlier discussion that cross ratio is invariant under projection; so, we may conclude that $(EF, GH) = (BD, IH)$ by a projection from the point A. However, $(BD, IH) = (FE, GH)$ through a projection from the point C; and thus we can see that $(EF, GH) = (FE, GH)$. But we observed earlier from R(2) that if the first pair of points of a cross ratio is interchanged, then the new cross ratio is the reciprocal of the first cross ratio. Consequently, if we let $(EF, GH) = y$, then from R(2) we get $(FE, GH) = 1/y$; but since we have shown that $(EF, GH) = (FE, GH)$, it follows that $y = 1/y$. Then, $y^2 = 1$, and $y = \pm 1$. But, since $(EF, GH) = (EG/GF)/(EH/HF)$, where EG, GF, EH are positive and HF is negative, (EF, GH) must be negative. Therefore, we can conclude that $(EF, GH) = -1$.

Concluding Comments

There are many branches of mathematics that have proved far more fruitful to the world of mathematics than projective geometry. For example, today topology is being richly developed by an enthusiastic group of researchers; and other geometric types of research, such as that being done in differential geometry, have proved more useful in varied applications. However, few mathematical scholars would dispute the fact that projective geometry has provided some of the most elegant, intriguing, and fascinating results in all of mathematics. As Kline so appropriately noted of projective geometry, "The science born of art proved to be an art" (1956, p. 641).

BIBLIOGRAPHY

Adler, Irving. *A New Look at Geometry*. New York: John Day, 1966.

Argunov, B. I., and L. A. Skornyakov. *Configuration Theorems*. Boston: D. C. Heath & Co., 1963.

Courant, Richard, and Herbert Robbins. *What Is Mathematics?* New York: Oxford University Press, 1941.

Coxeter, H. S. M. *Introduction to Geometry*. New York: John Wiley & Sons, 1961.

———. *Projective Geometry*. New York: Blaisdell Publishing Co., 1964.

Dorwart, Harold L. *The Geometry of Incidence*. Englewood Cliffs, N. J.: Prentice-Hall, 1966.

Dowdy, S. M. *Mathematics: Art and Science*. New York: John Wiley & Sons, 1971.

Eves, Howard. *A Survey of Geometry*. Rev. ed. Boston: Allyn & Bacon, 1972.

Gans, David. *Transformations and Geometries*. New York: Appleton-Century-Crofts, 1969.

Hewitt, Frances. "A New Look at Some Old Geometry Problems." In *Enrichment Mathematics for High School*. Twenty-eighth Yearbook of the National Council of Teachers of Mathematics, pp. 65–75. Washington, D.C.: The Council, 1963.

Hilbert, D., and S. Cohn-Vossen. *Geometry and the Imagination*. New York: Chelsea Publishing Co., 1952.

Holton, Jean Laity. *Geometry: A New Way of Looking at Space*. New York: Weybright & Talley, 1971.

Kline, Morris. *Mathematics in Western Culture*. New York: Oxford University Press, 1953.

———. "Projective Geometry." In *The World of Mathematics*, vol. 1. Edited by James R. Newman. New York: Simon & Schuster, 1956.

Kramer, Edna E. *The Nature and Growth of Modern Mathematics*. New York: Hawthorn Books, 1970.

Meserve, Bruce E. *Fundamental Concepts of Geometry*. Reading, Mass.: Addison-Wesley Publishing Co., 1955.

Pedoe, Daniel. *An Introduction to Projective Geometry*. New York: Macmillan Co., 1963.

Seidenberg, A. *Lectures in Projective Geometry*. New York: D. Van Nostrand Co., 1962.

Springer, C. E. *Geometry and Analysis of Projective Spaces*. San Francisco: W. H. Freeman & Co., 1964.

Veblen, Oswald, and John W. Young. *Projective Geometry*, vol. 1. New York: Blaisdell Publishing Co., 1938.

Young, John W. *Projective Geometry*. Carus Mathematical Monographs, no. 4. Washington, D.C.: Mathematical Association of America, 1930.

3

Groups

Roy Dubisch

What Is a Group?

In the high court of the kingdom of Lower Slobbovia, there are three judges for every trial. In the grand tradition of algebra, let's call them A, B, and C. They file in at the beginning of any trial and sit as shown in figure 3.1.

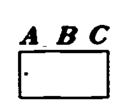

Fig. 3.1

But when the eccentric king of Lower Slobbovia (who attends all trials) yells "Promenade 1," B and C change places; when he yells "Promenade 2," A and B change places; and when he yells "Promenade 3," A goes to where C was sitting, B goes to where A was sitting, and C goes to where B was sitting. If the judges were sitting in their original positions when the promenade calls were given, the results would be as shown in figure 3.2.

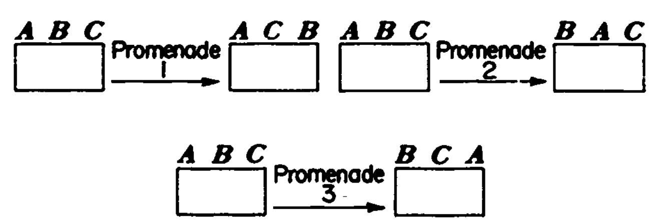

Fig. 3.2

Now in a hectic mood one day, the king yells "Promenade 1" and, two minutes later, yells "Promenade 2." We can picture the action as in figure 3.3.

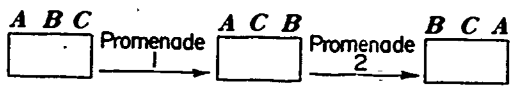

Fig. 3.3

To his royal amazement, the king realizes that the judges are now seated exactly as they would be if, instead of yelling "Promenade 1" and then "Promenade 2," he had just yelled "Promenade 3."

The next day he decides to try this procedure again—but with a slight variation: Now he yells "Promenade 2" first and then yells "Promenade 1." What happens is shown in figure 3.4.

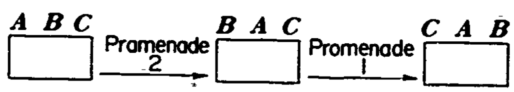

Fig. 3.4

Now he is amazed to find out that the result is not the same as Promenade 3; indeed, the result is what he has been calling Promenade 4.

At this point he decides to call in the court mathematician to analyze the whole business. Immediately, like any mathematician, the C.M. employs some notation. He lets P_1 stand for Promenade 1, P_2 for Promenade 2, and so forth, and writes

$$P_1 = \begin{pmatrix} A & B & C \\ A & C & B \end{pmatrix}, \quad P_2 = \begin{pmatrix} A & B & C \\ B & A & C \end{pmatrix}, \quad P_3 = \begin{pmatrix} A & B & C \\ B & C & A \end{pmatrix}, \quad P_4 = \begin{pmatrix} A & B & C \\ C & A & B \end{pmatrix}$$

Then he writes

$$P_1P_2 = \begin{pmatrix} A & B & C \\ A & C & B \end{pmatrix} \begin{pmatrix} A & B & C \\ B & A & C \end{pmatrix} = \begin{pmatrix} A & B & C \\ B & C & A \end{pmatrix} = P_3$$

by noting that, under P_1, A goes to A and then under P_2, A goes to B; that under P_1, B goes to C, and then under P_2, C goes to C; and that, under P_1, C goes to B, and then, under P_2, B goes to A. Symbolically,

$$\underbrace{A \xrightarrow{P_1} A \xrightarrow{P_2} B}_{P_3}; \quad \underbrace{B \xrightarrow{P_1} C \xrightarrow{P_2} C}_{P_3}; \quad \underbrace{C \xrightarrow{P_1} B \xrightarrow{P_2} A}_{P_3}.$$

Similarly,

$$P_2P_1 = \begin{pmatrix} A & B & C \\ B & A & C \end{pmatrix} \begin{pmatrix} A & B & C \\ A & C & B \end{pmatrix} = \begin{pmatrix} A & B & C \\ C & A & B \end{pmatrix} = P_4$$

and

$$P_2P_3 = \begin{pmatrix} A & B & C \\ B & A & C \end{pmatrix} \begin{pmatrix} A & B & C \\ B & C & A \end{pmatrix} = \begin{pmatrix} A & B & C \\ C & B & A \end{pmatrix},$$

a new promenade which he calls P_5, and

$$P_1P_1 = \begin{pmatrix} A & B & C \\ A & C & B \end{pmatrix} \begin{pmatrix} A & B & C \\ A & C & B \end{pmatrix} = \begin{pmatrix} A & B & C \\ A & B & C \end{pmatrix},$$

a non-Promenade which he calls I (for identity), since $P_1I = IP_1 = P_1$, $P_2I = IP_2 = P_2$, and so on.

He now decides to systematize his work by making a "multiplication" table:

	I	P_1	P_2	P_3	P_4	P_5
I	I	P_1	P_2	P_3	P_4	P_5
P_1	P_1	I	P_3	P_2	P_5	P_4
P_2	P_2	P_4	I	P_5	P_1	P_3
P_3	P_3	P_5	P_1	P_4	I	P_2
P_4	P_4	P_2	P_5	I	P_3	P_1
P_5	P_5	P_3	P_4	P_1	P_2	I

Exercise 1. Check his work!

As the C.M. studies the situation, he realizes quickly that he could work out a similar table for the changing around or *permutation* of four or more people. Thus, for example,

$$\begin{pmatrix} A & B & C & D \\ A & C & B & D \end{pmatrix} \begin{pmatrix} A & B & C & D \\ B & C & D & A \end{pmatrix} = \begin{pmatrix} A & B & C & D \\ B & D & C & A \end{pmatrix}$$

Also, he notes considerable similarity between the combination of permutations as shown in the "promenade" table and the multiplication of positive rational numbers—and one significant difference.

1. Just as $a \cdot 1 = 1 \cdot a = a$ for all positive rational numbers a, so, he notes, $PI = IP = P$ for all permutations P.

2. Just as every positive rational number a has a multiplicative inverse (i.e., a number b such that $ab = ba = 1$), so, he notes, for every permutation P there is a permutation Q such that $PQ = QP = I$. Examples: $P_1P_1 = I$; $P_2P_2 = I$; $P_5P_5 = I$ (i.e., P_1, P_2, and P_5 are their own inverses); $P_3P_4 = P_4P_3 = I$.

3. After noting that $P_1(P_2P_3) = P_1P_5 = P_4$ and $(P_1P_2)P_3 = P_3P_3 = P_4$, that $P_2(P_4P_3) = P_2I = P_2$ and $(P_2P_4)P_3 = P_1P_3 = P_2$, and that $P_5(P_3P_2) = P_5P_1 = P_3$ and $(P_5P_3)P_2 = P_1P_2 = P_3$, he concludes (correctly) that if P, Q, and R are any permutations, then $(PQ)R = P(QR)$. That is, like multiplication of positive rational numbers, combination of permutations is associative.

4. He notes, as did his majesty, that $P_1P_2 \neq P_2P_1$ and concludes that combination of permutations, unlike multiplication of positive rational numbers, is not a commutative operation.

Exercise 2. Find all pairs (P, Q) such that $PQ = QP$.

Exercise 3. Show that $P_1(P_4P_5) = (P_1P_4)P_5$ and that $P_2(P_3P_5) = (P_2P_3)P_5$.

The C.M. is well on his way to a study of what other mathematicians have, for over a hundred years, called *group theory*. Formally, a *group* is a set G with an operation defined on pairs of elements of G (i.e., a *binary* operation)

such that, if $*$ represents the operation, we have the following properties:

1. If a and b are in G, then $a * b$ is in G. (Closure property)
2. There exists an element e of G such that $a * e = e * a = a$ for all a in G. (Existence of identity)
3. For any a in G, there exists a b in G such that $a * b = b * a = e$. (Existence of inverses)
4. $a * (b * c) = (a * b) * c$ for all a, b, c in G. (Associative property)

When no confusion can arise, we usually do as the C.M. did and write ab for $a * b$.

Groups abound in mathematics. We have seen that we can have a group of permutations, and it is easy to see that the positive rational numbers form a group under multiplication. There are, however, two important differences between these two groups: (1) The permutation group has a finite number of elements and is an example of a *finite* group, whereas the set of positive rational numbers is an infinite set and so provides us with an example of an *infinite* group. (2) The operation of multiplication of rational numbers is a commutative operation (i.e., $ab = ba$ for all rational numbers a and b). Thus the group of positive rational numbers under multiplication is an example of a *commutative* group, whereas the permutation groups (except for those on 1 or 2 letters!) are examples of noncommutative groups.

Other examples of infinite commutative groups are the integers under addition and the nonzero real numbers under multiplication.

Exercise 4. Give some other examples of infinite commutative groups.

Exercise 5. Why does not the set of *all* real numbers form a group under multiplication?

(Our example of a noncommutative group was a finite group, whereas all of our examples of commutative groups were infinite. As we shall see, however, there are commutative finite groups. Noncommutative infinite groups also exist, but the description of them is rather technical.)

Isomorphic Groups

Now let us consider the set of *symmetries* of an equilateral triangle ABC, figure 3.5. We can rotate this triangle counterclockwise through 120° and it will look like figure 3.6. That is, it will appear unchanged except for the labels on the vertices. Let's call this rotation R. Similarly, a counterclockwise rotation through 240° will give us figure 3.7, and we'll call this

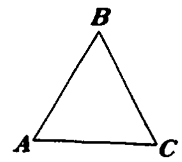

Fig. 3.5

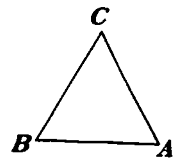

Fig. 3.6

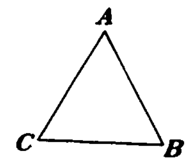

Fig. 3.7

rotation R'. A rotation through 360°, of course, will give us back the original triangle labeled as in figure 3.5, and we can consider this as the identity, I.

We can also "flip" the triangle around any of the medians to obtain F_1, F_2, and F_3 as shown in figure 3.8.

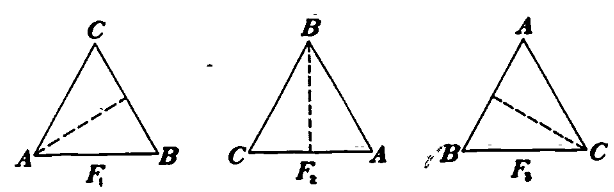

Fig. 3.8

Now we can talk about consecutive movements. See, for example, figure 3.9.

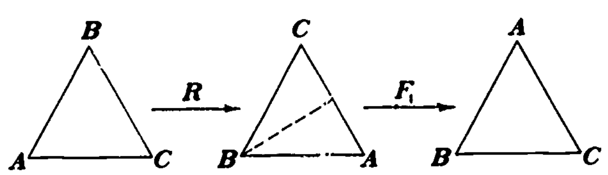

Fig. 3.9

But this is the same result as if we had just applied F_3 to the triangle in its original position, and so we write $RF_1 = F_3$. Now let's compute F_1R. We have the result shown in figure 3.10 and conclude that $F_1R = F_2$.

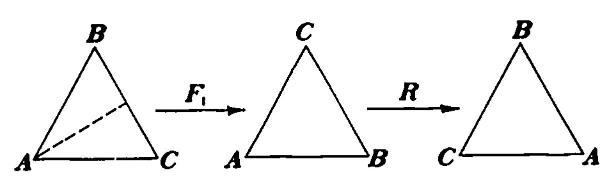

Fig. 3.10

Now we could continue in this way finding other "products" and establishing the fact that I, R, R', F_1, F_2, and F_3 form a group under the operation of triangle movements. An easier way of establishing this conclusion, however, is to observe that each of our triangle movements results in a permutation of the three vertices and that I, R, R', F_1, F_2, and F_3 account for all possible permutations of A, B, and C. Thus, using the notation that we employed before, we have

$$R \longleftrightarrow \begin{pmatrix} A & B & C \\ C & A & B \end{pmatrix} = P_4$$

(i.e., after the rotation R, A labels the vertex originally labeled C; B labels

the vertex originally labeled A; and C labels the vertex originally labeled B). Similarly,

$$R' \longleftrightarrow \begin{pmatrix} A & B & C \\ B & C & A \end{pmatrix} = P_3,$$

$$F_1 \longleftrightarrow \begin{pmatrix} A & B & C \\ A & C & B \end{pmatrix} = P_1,$$

$$F_2 \longleftrightarrow \begin{pmatrix} A & B & C \\ C & B & A \end{pmatrix} = P_5,$$

$$F_3 \longleftrightarrow \begin{pmatrix} A & B & C \\ B & A & C \end{pmatrix} = P_2,$$

and, of course,

$$I \longleftrightarrow \begin{pmatrix} A & B & C \\ A & B & C \end{pmatrix} = I.$$

Then $RF_1 = F_3$ corresponds to $P_4P_1 = P_2$, $F_1R = F_2$ corresponds to $P_1P_4 = P_5$, and so forth. That is, except for notation, the group of symmetries of an equilateral triangle is identical with the permutation group on three letters.

When two groups differ like this only in notation we say that the two groups are *isomorphic*. This concept of isomorphism occurs in many places in mathematics and is extremely important in that it enables us to show that some seemingly different systems are basically the same. (We use this idea in doing multiplication by the use of logarithms: the group of positive real numbers under multiplication is isomorphic to the additive group of all real numbers. The correspondence is $a \longleftrightarrow \log_{10}a$ for any positive real number a and we have $a \times b \longleftrightarrow \log_{10}(a \times b) = \log_{10}a + \log_{10}b$.)

Subgroups

From any group G we can extract *subgroups*—subsets of G which themselves form a group under the group operation of G. For example, the group of permutations on three letters has the subgroups

$\{I, P_3, P_4\}$

	I	P_3	P_4
I	I	P_3	P_4
P_3	P_3	P_4	I
P_4	P_4	I	P_3

$\{I, P_1\}$

	I	P_1
I	I	P_1
P_1	P_1	I

$\{I\}$

	I
I	I

Exercise 6. Find two other subgroups of the permutation group on three letters.

Any group G with identity e has the special subgroups $\{e\}$ and G itself. The search for other subgroups of *finite* groups is greatly facilitated by the use of *La Grange's* theorem, which states that *if S is a subgroup of a group G,*

then the order *of S (the number of elements of S) must divide the order of G.* Thus any subgroup of the permutation group on three letters must have order 1, 2, 3, or 6.

For infinite groups we have no such useful theorem; but it is easy to see, for example, that the group of all positive rational numbers under multiplication is a subgroup of the group of all positive real numbers under multiplication and that the group of all integers under addition is a subgroup of the group of all rational numbers under addition.

Exercise 7. Does the set of all positive irrational numbers form a subgroup of the group of all positive real numbers under multiplication?

Cyclic Groups

Of special importance is a class of groups known as *cyclic* groups. A cyclic group of order n can be described very simply as a group with elements $I, R, R^2, R^3, \cdots, R^{n-1}$ where $R^2 = RR$, $R^3 = RR^2$, $R^4 = RR^3$, $\cdots$, and $R^n = I$. Thus, for example, when $n = 3$, we have

	I	R	R^2
I	I	R	R^2
R	R	R^2	I
R^2	R^2	I	R

(We get $R^2R^2 = R$ by $R^2R^2 = (RR)R^2 = R(RR^2) = RR^3 = RI = R$.)

If you are familiar with clock (modular) arithmetic, you can easily see that this group is isomorphic to three clock arithmetic (arithmetic modulo 3) under the operation of addition. We have

+	0	1	2
0	0	1	2
1	1	2	0
2	2	0	1

and the isomorphism correspondence is given by $0 \longleftrightarrow I$, $1 \longleftrightarrow R$, and $2 \longleftrightarrow R^2$. Indeed, a cyclic group of order n is isomorphic to arithmetic modulo n under the operation of addition. (Note that these are examples of finite commutative groups.)

Every infinite cyclic group, on the other hand, is isomorphic to the group of integers under the operation of addition. We can write $I \longleftrightarrow 0$, $R \longleftrightarrow 1$, $R^2 \longleftrightarrow 2, \cdots, R^n \longleftrightarrow n$, $R^{-1} \longleftrightarrow -1$, $R^{-2} \longleftrightarrow -2, \cdots, R^{-n} \longleftrightarrow -n$ and write the cyclic group as $\{\cdots, R^{-2}, R^{-1}, I, R, R^2, \cdots\}$. Then, for example, just as $1 + 2 = 3$, so $RR^2 = R^3$ and just as $-2 + 2 = 0$, so $R^{-2}R^2 = I$.

Many books and articles have been and continue to be written on group theory. In particular, any book on abstract algebra will have at least one chapter on group theory. In the books listed at the end of the article you will find additional references that can provide for a lifetime of study of groups!

Answers to Exercises

1. Given.

2. Of course, $P_iI = IP_i$ for $i = 1, 2, 3, 4$, or 5 and if $P = Q$ we certainly have $PQ = QP$. Also, however, we have $P_3P_4 = P_4P_3$.

3. $P_1(P_4P_5) = P_1P_1 = I$, and $(P_1P_4)P_5 = P_5P_5 = I$; $P_2(P_3P_5) = P_2P_2 = I$, and $(P_2P_3)P_5 = P_5P_5 = I$.

4. Two examples are the rational numbers under addition and the nonzero rational numbers under multiplication.

5. 1 is the identity for multiplication, but there is no real number a such that $0 \times a = 1$.

6.

$$\begin{array}{c} \{I, P_2\} \\ \begin{array}{c|cc} & I & P_2 \\ \hline I & I & P_2 \\ P_2 & P_2 & I \end{array} \end{array} \qquad \begin{array}{c} \{I, P_5\} \\ \begin{array}{c|cc} & I & P_5 \\ \hline I & I & P_5 \\ P_5 & P_5 & I \end{array} \end{array}$$

(Recall that the group itself is also considered to be a subgroup of itself.)

7. No. We do not have closure, since, for example, $\sqrt{2}$ is an irrational number but $\sqrt{2}\,\sqrt{2} = 2$ is a rational number.

REFERENCES

Adler, Irving. *Groups in the New Mathematics.* New York: John Day Co., 1967.

Andree, Richard V. *Selections from Modern Abstract Algebra.* New York: Henry Holt & Co., 1958.

Burton, David M. *An Introduction to Abstract Mathematical Systems.* Reading, Mass.: Addison-Wesley Publishing Co., 1965.

Carmichael, R. D. *Introduction to the Theory of Groups of Finite Order.* New York: Dover Publications, 1956.

Dubisch, Roy. *Introduction to Abstract Algebra.* New York: John Wiley & Sons, 1965.

Grossman, Israel, and Wilhelm Magnus. *Groups and Their Graphs.* New York: Random House, 1964.

Lieber, L. R. *Galois and the Theory of Groups.* Lancaster, Pa.: Science Press, 1932.

Yarnelle, J. E. *Finite Mathematical Structures.* Lexington, Mass.: D. C. Heath & Co., 1964.

4

Infinity and Transfinite Numbers

Sister Conrad Monrad, S.P.

Classical Problems Involving Infinity

Infinity is a concept that boggles our minds and yet intrigues us. It is a difficult concept for us to comprehend, since nothing in the world about us seems to possess the property of being "infinite"—except possibly space or time. Even these, despite their appearance of being endless or limitless, are actually finite in the personal experience of any individual. Examples such as the grains of sand on the seashore, or the electrons and protons in the universe, constitute large, but finite, sets. For, although we are physically incapable of counting them, they are limited by some large number that is finite. Perhaps our best example of an infinite set is displayed in our experience with the counting numbers, or the natural numbers. As small children we found there was no limit or end to the natural numbers, that no matter how far you counted, there were more numbers that followed. If any one claims to name the last or largest counting number, we can easily name a larger number by adding "one" to his "largest number." The concept of infinity in mathematics has many aspects—the infinitely many, the infinitely large, the infinitely small, infinite divisibility, infinite repetition, infinite summation, and a great variety of infinite processes. Infinity is versatile and surprising.

As man moved from the use of mathematics for strictly pragmatic purposes in Babylonia and in Egypt to leisurely speculation about mathematics in

Greek civilization, it was the notion of infinity that "disturbed the peace." Zeno of Elea, a Greek philosopher of the fifth century before Christ, tossed an upsetting element into the active world of Greek thought by his four paradoxes involving two diametrically opposite viewpoints of infinity and motion. In his paradox of Achilles and the tortoise, Achilles, who is running, cannot overtake the crawling tortoise ahead of him because he must first reach the former position of the tortoise—but when Achilles reaches that place, the tortoise has moved on, and so is still ahead. As this is repeated each time, the tortoise will always be ahead of Achilles.

In 1632 Galileo presented in his *Dialogue on Two New Sciences* some perplexing questions on infinity to the thinkers of his century. One of these was concerned with the application of words such as "equal," "greater," and "less" to infinite quantities. He compared the apparently greater infinity of points contained in a long line segment with the smaller infinity of points contained in a shorter line segment. Then he noticed that the numbers that are squares of the natural numbers are a subset of the natural numbers and that the natural numbers have more numbers than the squares only. Yet, by matching each number with its square, Galileo showed that there are as many squares as there are natural numbers.

$$\begin{array}{cccccccc} 1 & 2 & 3 & 4 & 5 & 6 & \ldots \\ 1 & 4 & 9 & 16 & 25 & 36 & \ldots \end{array}$$

From this he concluded that any two infinite sets have a corresponding number of elements and, as a result, one line segment did not contain "more," "less," or "as many" points as another, but that each contained an infinity of elements. The question implied by his conclusions, that is, whether the infinity of points in a line segment is equivalent to the infinity of squares of natural numbers, was not settled until 1873.

Georg Cantor, two hundred years later, started where Galileo left off. He devised a method of displaying a one-to-one correspondence between the natural numbers and the rational numbers, as well as between any two line segments. However, he questioned the possibility of pairing the natural numbers with the real numbers. He felt that their nature prevented such a pairing since the natural numbers "consist of discrete parts" while the real numbers form a "continuum." At first, however, he could find no supporting reason. In 1873 he found his first proof and in 1890 a second, simpler proof. Thus, he resolved Galileo's problem.

Cantor, in turn, raised a question that was not answered for another century. "Is every infinite subset of the reals either equivalent to the reals, or to the natural numbers?" The continuum hypothesis is the statement of Cantor's belief that the answer to this question is yes. It was not until 1963 that Paul J. Cohen settled this question by proving that the continuum hypothesis is not a consequence of the commonly accepted fundamental properties of sets. Finally, it was Bertrand Russell who solved Zeno's paradox of Achilles and the tortoise by using Cantor's conclusion that any two line segments have the same number of points.

Denumerable Sets

In more recent times David Hilbert is credited with inventing a paradox called "Hilbert's Hotel," similar to Galileo's paradox of the squares of positive integers. As the story goes, a guest comes to Hilbert's Hotel and requests a room for the night, only to be told that the hotel is full. But the manager then proceeds to solve the problem by putting the guest in room 1, moving the occupant of room 1 to room 2, the occupant of room 2 to room 3, and, in general, the occupant of room n to room $n + 1$. The hotel had an infinite number of rooms! How, then, could the manager claim that the hotel was full?

While Hilbert's Hotel is a fantasy and not a real building, the world of mathematics contains an abundance of infinite sets or collections. For example, the set of all natural numbers or counting numbers: 1, 2, 3, 4, 5, . . . ; the set of all positive even integers: 2, 4, 6, 8, 10, . . . ; the set of all primes: 2, 3, 5, 7, 11, . . . ; the set of all positive multiples of 5: 5, 10, 15, 20, 25, . . . ; the set of all squares of positive integers: 1, 4, 9, 16, 25, . . . ; the set of all representations of 1/3 in the form of ratios of integers: 1/3, 2/6, 3/9, 4/12, 5/15, . . . ; the set of all reciprocals of the positive integers: 1, 1/2, 1/3, 1/4, 1/5, . . . , are all infinite sets. Some other examples of infinite sets are the set of points on a straight line, the set of circles in a plane, the set of translations in a plane, and the set of cubes in space.

Cantor compared infinite sets by setting up a one-to-one correspondence between their elements, that is, by matching exactly one element of either set to exactly one element of the other set, until all matched. We do this with finite sets—the places at the dinner table are matched with the members of the family, and the visitors with the set of empty chairs in the room. Two sets are said to be equivalent if there is a one-to-one correspondence between them. We can show that several of the sets mentioned above are equivalent to the set of natural numbers by setting up a correspondence of each set to the natural numbers as follows:

the set of even positive integers,

$$\begin{array}{l} 1,\ 2,\ 3,\ 4,\ 5,\ \ldots,\ n,\ \ldots \\ 2,\ 4,\ 6,\ 8,\ 10,\ \ldots,\ 2n,\ \ldots \end{array}$$

the set of positive multiples of 5,

$$\begin{array}{l} 1,\ 2,\ 3,\ 4,\ 5,\ \ldots,\ n,\ \ldots \\ 5,\ 10,\ 15,\ 20,\ 25,\ \ldots,\ 5n,\ \ldots \end{array}$$

the set of reciprocals of the positive integers,

$$\begin{array}{l} 1,\ 2,\ 3,\ 4,\ 5,\ \ldots,\ n,\ \ldots \\ 1,\ 1/2,\ 1/3,\ 1/4,\ 1/5,\ \ldots,\ 1/n,\ \ldots \end{array}$$

the set of representatives of 1/3,

$$\begin{array}{l} 1,\ 2,\ 3,\ 4,\ 5,\ \ldots,\ n,\ \ldots \\ 1/3,\ 2/6,\ 3/9,\ 4/12,\ 5/15,\ \ldots,\ n/3n,\ \ldots \end{array}$$

the set of primes,

$$\begin{array}{l} 1,\ 2,\ 3,\ 4,\ 5,\ \ldots \\ 2,\ 3,\ 5,\ 7,\ 11,\ \ldots \end{array}$$

Here, corresponding to n is the nth smallest prime. Any set that is equivalent to the set of natural numbers is a denumerable set and is denumerably, or countably, infinite.

Infinity is full of surprises. The set of odd natural numbers is an infinite subset of the natural numbers. We would expect the set of natural numbers to have more members than its subset of odd numbers because the natural numbers contain the even numbers as well as the odd numbers. If to each natural number n we assign the odd number $2n - 1$,

$$\begin{array}{l} 1,\ 2,\ 3,\ 4,\ 5,\ \ldots,\ n,\ \ldots \\ 1,\ 3,\ 5,\ 7,\ 9,\ \ldots,\ 2n-1,\ \ldots \end{array}$$

then there is an odd number corresponding to each natural number, and, behold, the subset of odd numbers is equivalent to the set of natural numbers. This means the set of odd natural numbers is a denumerable set. Georg Cantor used a generalization of this last example to characterize an infinite set. He said that any infinite set is equivalent to a proper subset of itself (a "proper" subset does not contain *all* the members of the set). There is no finite set that matches up one-to-one with a proper subset of itself. Consider the finite set $M = \{2, 4, 6, 8, 10\}$. No subset of M, other than M itself, can be put into a one-to-one correspondence with M.

Suppose S is the set of all negative integers and the reciprocals of all positive integers. No doubt, one could establish a correspondence of the negative integers to the natural numbers, or of the reciprocals of the positive integers, but is S itself equivalent to the natural numbers? Is S countably infinite? First, arrange S so that $-n$ precedes $1/n$:

$$-1,\ 1,\ -2,\ 1/2,\ -3,\ 1/3,\ -4,\ 1/4,\ \ldots,\ -n,\ 1/n,\ \ldots$$

To each odd natural number, n_o, match $\frac{-(n_o + 1)}{2}$, and to each even natural number, n_e, match $\frac{2}{n_e}$:

$$\begin{array}{llllllll} 1, & 2, & 3, & 4, & 5, \ldots, & & n_o, & n_e, \ldots \\ -1, & 1, & -2, & 1/2, & -3, \ldots & & \frac{-(n_o+1)}{2}, & \frac{2}{n_e}, \ldots \end{array}$$

Then S must be a denumerable set. One c see from these few examples that an abundance of denumerable sets exists.

Rationals as a Denumerable Set

Are the rational numbers a denumerable set? This question is not so simply answered. When we look at the rational numbers, their density poses a problem. For any two rational numbers—no matter how small their difference—there is always a rational number between them. How can the rationals be lined up, then, in order to set up some type of correspondence? Is it possible? As a stud... Cantor devised a method of doing this. The following describes one such arrangement of the rationals.

We can begin by assigning a value to each rational number equal to the sum of its denominator and the absolute value of its numerator, and group together all those with the same value. After eliminating repetitions, like 4/2 and 6/3, and any rationals not in simplest form, that is, any with common factors (greater than 1) in numerator and denominator, the groups can be arranged according to their value. Only one rational, 0/1, has value 1, and two, −1/1 and 1/1, have value 2. Within each group the numbers can be arranged by pairs, the negative number first, in ascending order according to the value of the positive number. For example, the group of value 4 would be arranged as follows: −1/3, 1/3, −3/1, 3/1. We have lined up all the groups now by their values, 1, 2, 3, 4, . . . , and each number within its group. All one needs to do is to assign a natural number to each to have a one-to-one correspondence. The rational numbers are a denumerable set!

1,	2,	3,	4,	5,	6,	7,	8,	9,	. . .
0/1,	−1,	1,	−1/2,	1/2,	−2,	2,	−1/3,	1/3,	. . .

Are the Real Numbers Denumerable?

The investigation of infinite sets seems to lead to the same conclusion that Galileo reached—that all infinite sets are simply infinite and cannot be classified in any other way. Before stopping, however, examine the real numbers. Note that a real number may be written as a nonterminating decimal, and every nonterminating decimal is a real number. Assuming that there is a one-to-one correspondence between the real numbers and the natural numbers, let each real number be represented by a nonterminating decimal in this correspondence. If the following represents the correspondence, where the a's represent digits in the nonterminating decimals, then one can find a nonterminating decimal, that is, a real number, that is not given in this listing—this, despite the fact that all of them had been listed!

$$\begin{array}{ll}
1 & 0.a_{11}\,a_{12}\,a_{13}\,a_{14}\,a_{15}\ldots \\
2 & 0.a_{21}\,a_{22}\,a_{23}\,a_{24}\,a_{25}\ldots \\
3 & 0.a_{31}\,a_{32}\,a_{33}\,a_{34}\,a_{35}\ldots \\
4 & 0.a_{41}\,a_{42}\,a_{43}\,a_{44}\,a_{45}\ldots \\
\cdot & \\
\cdot & \\
\cdot &
\end{array}$$

Replace each digit along the diagonal, a_{11}, a_{22}, a_{33}, . . ., by a different one, b_1, b_2, b_3, . . ., where b_n is not 0 or 9. The number given by the nonterminating decimal, $0.b_1b_2b_3\ldots$, is a real number not contained in our list, since it differs from each given number by at least one digit. In other words, if ever there was found a way of setting up a correspondence matching up *all* the real numbers with the natural numbers, one could at once name a real number that had been omitted! Since this is always possible, it means one cannot set up a one-to-one correspondence of the reals to the natural numbers—that no such correspondence exists. But this indicates that there is a difference between infinite sets, and that some are "greater" than others. Not all infinite sets then are

denumerably infinite: some, like the reals, are nondenumerable, or not countable.

The set of points on a line can be put into a one-to-one correspondence with the real numbers. Therefore, there are as many points on a line as there are real numbers. As a consequence the set of points on a line is a nondenumerable set. It follows that the points on a line are not equivalent to the natural numbers; therefore, Galileo's implied question has to be answered in the negative.

Cantor's work with infinite sets led many to believe that the properties of infinite sets could be used to distinguish one-space from two-space. This was not a result; on the contrary, the set of points on a line segment can be shown to be equivalent to the set of points in a square. The following theorem (Schröder-Bernstein theorem) is needed for this purpose: *If there is a one-to-one correspondence between a set A and a subset of B, and also a one-to-one correspondence between B and a subset of A, then there is a one-to-one correspondence between the sets A and B.* (See fig. 4.1.)

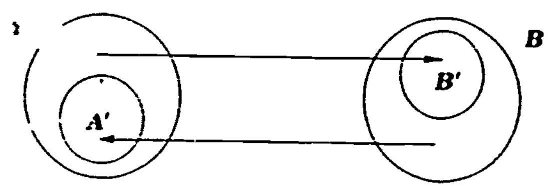

Fig. 4.1

If A is the unit segment and B is the unit square, then each point of A corresponds to a point of a subset of B, which could be a side of a square (fig. 4.2). Next, each point of the square B must be matched with a point of a subset of the segment A. A point in the square has real numbers, (a, b), as coordinates. These real numbers can be expressed as nonterminating decimals:

$$a = a_0.a_1a_2a_3 \ldots$$
$$b = b_0.b_1b_2b_3 \ldots .$$

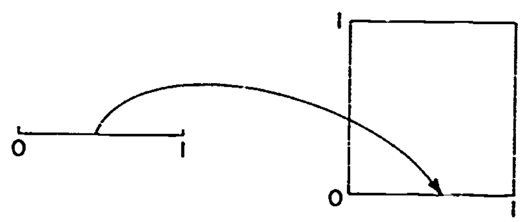

Fig. 4.2

Also, every point on the segment can be represented by a nonterminating decimal. By alternating the digits of a and b, another nonterminating decimal c is formed:

$$c = 0.a_0b_0a_1b_1a_2b_2 \ldots .$$

The new decimal is a real number between 0 and 1 and represents a point on the unit segment. (See fig. 4.3.)

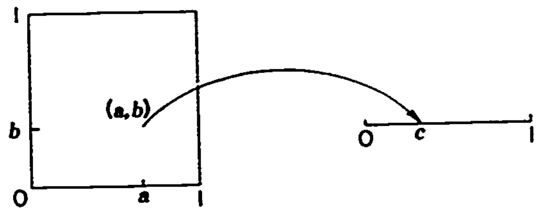

Fig. 4.3

Every point in the square then can be matched uniquely with a point on the segment. Since the conditions of the theorem are satisfied, the set of points on the segment is equivalent to the set of points in the square. Both are nondenumerable infinite sets, and it can be seen that infinity cannot be used to distinguish one-space from two-space.

Cardinality of Infinite Sets

In finite sets the cardinal number of a set indicates "how many" elements the set contains. For example, the cardinal number of the set of letters in the alphabet is 26. Our fingers form a set having the cardinal number 10. The cardinal number 10 is less than the cardinal number 26 because a set of 10 elements has fewer elements than a set of 26 elements. Cantor used the Hebrew letter $\aleph_0$, (read aleph null) for the cardinal number of all sets equivalent to the set of natural numbers. He designated c (for continuum) as the cardinal number of the reals. The cardinal number of an infinite set is called a *transfinite* number. $\aleph_0$ and c are transfinite numbers.

It appears that any infinite subset of the rationals or the natural numbers, such as the reciprocals of the positive integers or the even natural numbers, has cardinal number $\aleph_0$, that is, the *smallest* transfinite number must be $\aleph_0$! It is also evident from our work with the natural numbers and the reals that $\aleph_0 < c$. If there is no transfinite number smaller than $\aleph_0$, is there then one between $\aleph_0$ and c? Or, phrased another way, is $\aleph_1 = c$? Cantor intuitively responded in the affirmative—this is known as the continuum hypothesis—but he was not able to prove it. Recently, in 1940 Gödel, and in 1963 Cohen, showed the continuum hypothesis to be independent and in a position similar to that of the parallel postulate in geometry. That is, if the continuum hypothesis is added as a postulate to the accepted fundamental properties of set theory, you have a consistent axiomatic system comparable to Euclidean geometry. If, instead, a negation of the continuum hypothesis is added, one has a different, but consistent, system, as in non-Euclidean geometry.

The next question asked ought to be "Is there a transfinite number greater than c?" Before attempting to answer this question, first look at the use of exponents, as in $\aleph_0^2$ and $2^{\aleph_0}$. In finite cardinal numbers 3^2 means (3)(3),

or three threes. This suggests $\aleph_0^2 = (\aleph_0)(\aleph_0)$, or an infinite number of infinite sets. The set of ordered pairs of the positive integers would be an example of such a set:

```
(1,1) (1,2) (1,3) (1,4) ...
(2,1) (2,2) (2,3) (2,4) ...
(3,1) (3,2) (3,3) (3,4) ...
(4,1) (4,2) (4,3) (4,4) ...
  .
  .
  .
```

There are as many pairs in each row as there are natural numbers, and as many rows as there are natural numbers. A one-to-one correspondence can be set up as follows between each natural number in the following array and the ordered pair that is in the corresponding position in the array above:

```
 1 2 4 7 ...
  / / /
 3 5 8 ....
  / /
 6 9 .....
  /
10 ......
 .
 .
 .
```

But the cardinal number of the natural numbers is $\aleph_0$. This implies that $\aleph_0^2 = (\aleph_0)(\aleph_0) = \aleph_0$!

Next, examine the "power set" of a set S: this is the set of all possible subsets of S, including the empty set, ϕ, and S itself. For a set S of three elements, a, b, c, there are eight subsets: $S, \phi, \{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}$. For a set of four elements, there are sixteen subsets. Continue to increase the number in the set, and notice this:

For a set of three elements, there are $8 = 2^3$ subsets;
for a set of four elements, there are $16 = 2^4$ subsets; and
for a set of n elements, there are 2^n subsets.

So, for an infinite set of cardinality $\aleph_0$, the set of all possible subsets, that is, the power set of this infinite set, must have cardinality $2^{\aleph_0}$! How are n and 2^n related? We note that $3 < 2^3$, $4 < 2^4$, and, in general, $n < 2^n$. Then, $\aleph_0 < 2^{\aleph_0}$. Cantor showed further that $2^{\aleph_0} = c$. Using the same reasoning, $c < 2^c$, and behold, here is a transfinite number greater than c! Similarly, $2^c < 2^{(2^c)}$. As this process is repeated, another conclusion looms. There is no largest transfinite number!

Arithmetic with transfinite numbers is intriguing and full of surprises. For example, what is $\aleph_0 + \aleph_0$? Or $2\aleph_0$? What is the cardinality of the even natural numbers? Of the odd natural numbers? What set is the union of the even numbers and the odds? What is the cardinality of the natural numbers? Does this suggest that $\aleph_0 + \aleph_0 = \aleph_0$? Which of the operations and their properties that hold for finite cardinal numbers can be extended to transfinite numbers? Explore this further and see what you discover!

BIBLIOGRAPHY

Albrecht, Walter A., Jr. "Some Mathematical Concepts of Infinity." *Mathematical Log* 8 (April 1964): 1–3.

Berlinghoff, William P. *Mathematics, the Art of Reason*, pp. 180–204. Lexington, Mass.: D. C. Heath & Co., 1968.

Cantor, Georg. *Contributions to the Founding of the Theory of Transfinite Numbers.* Translated by P. E. B. Jourdain. LaSalle, Ill.: Open Court, 1915.

Courant, Richard, and Herbert Robbins. *What Is Mathematics?* pp. 77–86. New York: Oxford University Press, 1953.

Dantzig, Tobias. *Number, the Language of Science*, pp. 206–29. New York: Macmillan Co., 1930.

Douglis, Avron. *Ideas in Mathematics*, pp. 588–609. Philadelphia: W. B. Saunders Co., 1970.

Fraenkel, Abraham A. *Abstract Set Theory.* Amsterdam: North-Holland Publishing Co., 1953.

Fraleigh, John B. *Mainstreams of Mathematics*, pp. 58–97. Reading, Mass.: Addison-Wesley Publishing Co., 1969.

Gamow, George. *One, Two, Three, . . . Infinity*, pp. 14–23. New York: Viking Press, 1948.

Kamke, E. *Theory of Sets.* Translated by Frederick Bagemihl. New York: Dover Publications, 1950.

Kline, Morris. *Mathematics in Western Culture*, pp. 395–409. New York: Oxford University Press, 1953.

Meschkowski, Herbert. *Ways of Thought of Great Mathematicians*, pp. 91–104. Translated by John Dyer-Bennet. San Francisco: Holden-Day, 1964.

Newman, James R. *The World of Mathematics*, vol 3, pp. 1576–1611. New York: Simon & Schuster, 1956.

Quine, W. V. "Paradox." In *Mathematics in the Modern World, Readings from "Scientific American,"* pp. 97, 200–208, 212–20. Introductions by Morris Kline. San Francisco: W. H Freeman & Co., 1968.

Stein, Sherman K. *Mathematics, the Man-made Universe*, pp. 314–37. 2d ed. San Francisco: W. H. Freeman & Co., 1969.

Wilder, Raymond L. *Introduction to the Foundations of Mathematics*, pp. 78–107. New York: John Wiley & Sons, 1956.

Yarnelle, John E. *An Introduction to Transfinite Mathematics.* Lexington, Mass.: D. C. Heath & Co., 1964.

Zippin, Leo. *Uses of Infinity.* New York: L. W. Singer Co. 1962.

5

Pascal's Triangle

John D. Neff

There would seem to be an almost inexhaustible supply of problems and conjectures in the array commonly known as Pascal's triangle. One could allude to it either as a gold mine or as an iceberg—the former because the riches are there, but some ingenious labor is often needed; the latter because we shall perhaps never see more than a small percentage of the mass. Much of what you can learn will come through self-discovery, so you are encouraged to guess and experiment with easy cases at first and gradually pose conjectures and prove theorems later. In a few cases, the proof of the theorem will be very difficult, and you will have to settle for some of them in the references at the end.

The triangle appears in many different contexts at nearly all levels of mathematical endeavor. This is the real beauty of the triangle! The organization of the material that follows is as you might encounter the triangle in formal courses. As a suggestion for a talk, you might begin with a topic from the arithmetic section and then add material from any other section that interests you.

In its most familiar form the triangle appears as follows, with the row numbers appended for later reference.

Row	
0	1
1	1 1
2	1 2 1
3	1 3 3 1
4	1 4 6 4 1
5	1 5 10 10 5 1
6	1 6 15 20 15 6 1
7	1 7 21 35 35 21 7 1
8	1 8 28 56 70 56 28 8 1

The pattern, of course, is that each entry is the sum of the two entries immediately above it and each end entry is always the number 1. It is suggested

that for your talk you should always have the above nine rows in view. You might also save room for the next two rows, if it becomes necessary to display them.

Arithmetic

1. The row sum is always an even number. Once you are satisfied why this must be so, then show that the row sum in the nth row is 2^n.

2. In any row, the algebraic sum of the first minus the second plus the third minus the fourth, and so on, until the end, is always zero. Moreover, if you will watch carefully while doing this operation, you will find the numerical value of the partial sum conveniently nearby, to the northeast.

3. The sum of the first, third, fifth, seventh, etc., entries is the same as the sum of the second, fourth, sixth, eighth, etc., entries in any row. Why should it be half of the sum found in (1) above?

4. Number the diagonals of the triangle $d = 1, 2, 3, \ldots$ for later reference. (For example, the third diagonal contains the entries 1, 3, 6, 10, 15, ...) On any diagonal, the partial sum of the entries is conveniently nearby. (See Yaglom and Yaglom.)

5. Add the squares of the entries in any row and note that it is an entry in a later row. Describe the location of this sum on the squares in terms of the original row number used.

6. If you are very careful to "carry" over the digits, the nth row of the triangle is the number 11^n in disguise.

7. Choose any row (call it row A) and underline the far left entry. Choose a second row (call it row B) and underline any entry. Multiply these two entries and add to it the product of the corresponding entries moving to the right on row A and left on row B, until you run out of numbers. The sum of these products is in the triangle and can be described in terms of your starting position.

8. Pick any row of the triangle at random. Skip the first entry 1 entirely and then form the algebraic sum of the second entry, minus one-half of the third entry, plus one-third of the fourth entry, minus one-fourth of the fifth entry, and so on, until you reach the end of the row. This sum is the same as the partial sum of the "harmonic series" $1 + 1/2 + 1/3 + \ldots$, provided one can describe how many terms of the harmonic series are needed in terms of the row number chosen in the triangle. Incidentally, the harmonic series does not have a sum if the terms are added indefinitely.

Set Theory

9. How many subsets of k elements can be formed from a set with n distinct elements? For example, a set with two distinct elements has four subsets: the empty (or null) set, two single-element sets, and the original set itself. List all of the subsets similarly for other values of n, with $k = 1, 2, 3, \ldots, n$, until the pattern is clear. Don't forget that the null set is regarded as a subset of every set!

10. The symbol $\binom{n}{k}$ can be interpreted as the number of subsets with k elements that can be formed from a set with n distinct elements, and this symbol is assigned the numerical value

$$\frac{n!}{k!\,(n-k)!} = \binom{n}{k}.$$

(The symbol $0!$ is defined to have the value 1.) Form a triangle by considering the various values of k ($k = 0, 1, 2, 3, \ldots, n$) for each value of n, starting with $n = 0$. Each row number of this triangle will be the value of n selected and the entries arranged in increasing values of k. The first few rows will look like this:

Row					
0			$\binom{0}{0}$		
1		$\binom{1}{0}$		$\binom{1}{1}$	
2	$\binom{2}{0}$		$\binom{2}{1}$		$\binom{2}{2}$

Evaluate each of these symbols and continue adding rows to this triangle until the pattern is clear.

11. From a set containing $n + 1$ distinct elements $A, B, C, D, \ldots$, consider the subsets containing exactly r elements. There will be $\binom{n+1}{r}$ such subsets. Some of these subsets will contain a particular element (say, B) and the rest will not. Justify the "Pascal relation"

$$\binom{n+1}{r} = \binom{n}{r} + \binom{n}{r-1}$$

in terms of the number of subsets that contain the element B and those that do not contain the element B.

Algebra

12. One of the most widely known displays of the triangle is in connection with the successive positive integral powers of the binomial $(a + b)$. For example,

$$(a + b)^0 = 1;$$
$$(a + b)^1 = a + b;$$
$$(a + b)^2 = a^2 + 2ab + b^2.$$

Expand the product $(a + b)^n$ for enough more values of n until the pattern is clear.

13. In its compact form, the binomial theorem in (12) can be written

$$(a + b)^n = \sum_{k=0}^{n} \binom{n}{k} a^{n-k} b^k,$$

where the symbol $\binom{n}{k}$ is defined as in (10). Verify this form of the theorem, again using the convention $0! = 1$.

14. Clever choices of the numbers a and b in (12) or (13) will net you an easy proof of the statements made in (1) and (2), as well as a compact form for many of the statements made earlier. For example, the statement made in (5) can be written

$$\sum_{k=0}^{n} \binom{n}{k}^2 = \binom{2n}{n}$$

in this compact notation.

Plane Geometry

15. Place n points in a plane, so that no three are collinear, and connect this set of points with straight lines. How many lines have you drawn? Relate this answer to the triangle.

16. All of the lines that are drawn in (15) created an n-sided polygon and its diagonals. How many diagonals does a polygon with n sides have? (For example, an octagon has 20 diagonals.) Relate the number of diagonals in an n-sided polygon to the nth row of the triangle.

17. Place n points at random on a circle, and connect all of these points with straight lines (Bryant; Yaglom and Yaglom).

a) How many regions are formed inside the circle? You are creating the sequence 2, 4, 8, . . ., x, . . . starting with $n = 2$ points. Find the number x in terms of the triangle.

b) How many regions are formed inside the polygon that is inscribed in the circle? You are creating the sequence 1, 4, 11, . . ., y, . . . starting with $n = 3$ points. Find the number y in terms of the triangle.

c) Count the total number of intersection points of the diagonals of the polygon. You are creating the sequence 1, 5, 15, . . ., z, . . ., starting with $n = 4$ points. Find the number z in terms of the triangle.

18. This is much the same as the preceding, except that we shall use circles instead of lines. We wish to determine the maximum number of regions created in a plane by n circles, so drawn that no two are tangent, none is wholly inside or outside of another, and no three are concurrent. You are creating the sequence 2, 4, 8, 14, . . ., w, starting with $n = 1$ circle. (For example, with $n = 1$ circle, you have a region inside the circle and the region in the plane outside of the circle.) Find the number w in terms of the triangle. You will find it easier to relate the answer to the triangle by also considering the sequence formed by one-half of each number w (Bryant; Yaglom and Yaglom).

Probability

19. Continue the tree diagram in figure 5.1, which shows the two outcomes (H or T) of a fair coin tossed once and the four outcomes (HH, HT, TH, TT) of a fair coin tossed twice to the eight outcomes of three tosses, and so on. Count the number of ways that exactly k heads show up when the coin is

tossed n times, for several n and $k = 0, 1, 2, \ldots, n$. Considering each of the 2^n outcomes as equally likely when the coin is tossed n times, determine the probability of no heads, 1 head, 2 heads, . . ., n heads. Relate the numerator and the denominator of the probability to the triangle's nth row (Mosteller, Rourke, and Thomas).

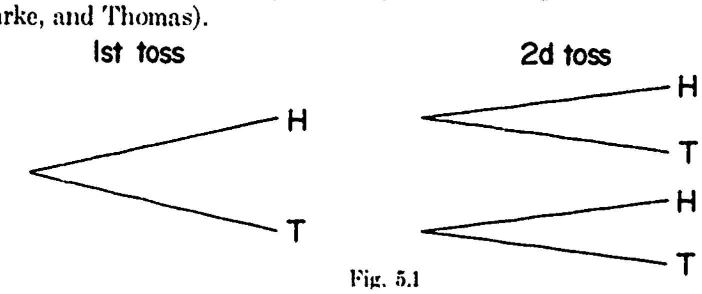

Fig. 5.1

20. How likely, in your opinion, would it be that both persons would get exactly the same number of heads when they each tossed a coin 10 times? The answer is perhaps surprising: more than one-sixth of the time this will happen (exactly .1872). Try to establish a pattern in the triangle by first considering that each person flips once, then twice, and so on and referring to the trees drawn in (19). In order to tackle the general case, you will need to refer back to (5) or (14) for the sum of the sum of the squares of the row entries. (The approximation $1/\sqrt{\pi n}$ is surprisingly good for large n. See Feller.)

Trigonometry

21. Verify the following trigonometric identities:

$$\cos 2\alpha = \cos^2\alpha - \sin^2\alpha,$$
$$\cos 3\alpha = \cos^3\alpha - 3\cos\alpha\sin^2\alpha, \text{ and}$$
$$\cos 4\alpha = \cos^4\alpha - 6\cos^2\alpha\sin^2\alpha + \sin^4\alpha.$$

and try to establish, with the aid of the nth row of the triangle, a similar identity for $\cos n\alpha$.

22. Verify the following trigonometric identities:

$$\sin 2\alpha = 2\sin\alpha\cos\alpha,$$
$$\sin 3\alpha = 3\sin\alpha\cos^2\alpha - \sin^3\alpha, \text{ and}$$
$$\sin 4\alpha = 4\sin\alpha\cos^3\alpha - 4\sin^3\alpha\cos\alpha,$$

and try to establish, with the aid of the nth row of the triangle, a similar identity for $\sin n\alpha$. The general pattern in this and the preceding problem is given by De Moivre's theorem, which can be found in any trigonometry book.

Solid Geometry

23. We would like to determine the maximum number of "chunks" of space that are created by n planes in arbitrary position. One plane creates 2 "chunks" (above and below), two planes can create a maximum of 4 "chunks" (3 if the planes are parallel and 2 if they are coincident). Convince yourself

that three planes will create at most 8 "chunks" by considering your classroom front wall, side wall, and floor. Four planes do not, unfortunately, create 16 "chunks," so try the easier two-dimensional problem "How many regions in a plane are created by n lines in arbitrary position?" You create the sequence 2, 4, 7, . . ., as you will quickly discover, starting with $n = 1$ line. Note that, to achieve the maximum, every new line you draw must cross every line already drawn in a distinct intersection point. Now consider the number of line segments formed by placing n points on a line. You can finish the original "chunks" problem and relate it to the triangle's nth row if you will finish table 5.1 and discover the pattern. (See Polya entries.)

TABLE 5.1

n	Number of Segments by n Points	Number of Regions by n Lines	Number of Chunks by n Planes
0	1	1	1
1	2	2	2
2	3	4	4
3	4	7	8

Calculus

24. The geometric series

$$1 + x + x^2 + x^3 + \dots = \frac{1}{1 - x}$$

for $|x| < 1$. (This can be easily verified by actual division.) Consider the corresponding derivatives

$$1 + 2x + 3x^2 + 4x^3 + \dots = \frac{1}{(1 - x)^2},$$

also a true statement for $|x| < 1$. By repeated differentiation, establish the connection between the successive powers of $\frac{1}{1 - x}$ and the numerical coefficients in the series of derivatives in terms of the triangle.

25. The decomposition of the rational function

$$\frac{n!}{x(x + 1)(x + 2) \dots (x + n)}$$

into partial fractions will involve the triangle very quickly. For example,

$$\frac{1}{x(x + 1)} = \frac{1}{x} - \frac{1}{x + 1}$$

and

$$\frac{2}{x(x + 1)(x + 2)} = \frac{1}{x} - \frac{2}{x + 1} + \frac{1}{x + 2}.$$

Try the cases $n = 3, 4, \dots$ in turn and observe the pattern. A similar situation exists for the rational function

$$\frac{n!}{x(x - 1)(x - 2)(x - 3) \dots (x - n)}$$

in terms of the nth row of the triangle.

Linear Algebra

26. Consider the matrix formed by the first two numbers in the first two diagonals, namely,

$$A = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}.$$

The determinant of this matrix has the value 1. Now write the matrix with elements the first three numbers in the first three diagonals, namely,

$$B = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \\ 1 & 3 & 6 \end{pmatrix},$$

and show that $|B| = 1$ also. Does the determinant of all such matrices formed in this manner always equal 1? If so, try to prove it.

Miscellaneous

If the iceberg analogy that was mentioned in the beginning is at all correct, this portion could be many pages long. However, the list of topics will close with some additional suggestions that are not as easily classified but are at least as interesting.

27. By drawing lines of approximately 30 degrees elevation through the triangle and adding the numbers touched by the lines, one will create the sequence 1, 1, 2, 3, 5, 8, 13, . . . known as the Fibonacci sequence. This sequence has many fascinating properties and applications, but only one will be mentioned here. The limit of the sequence formed by dividing each element by its successor is known as the golden section (approximate value 0.618 . . .). (See Hoggatt.)

28. Reconsider the tree drawn in (19) and ask for the probability that the first head occurs on the kth toss of the fair coin ($k = 1, 2, 3, \ldots$). The probability that the first head occurs on the first toss is one-half, since we are using a fair coin. This statement is written compactly as $P(k = 1) = 1/2$. Similarly, $P(k = 2) = 1/4$, since one needs a tail first and then a head in order to have the first head occur on the second trial. Is it possible that the first head will never occur? The answer is no! You need to sum the probabilities involved with the aid of the geometric series sum found in (24). In short, sooner or later, you must get the first head (Mosteller, Rourke, and Thomas).

29. Let us extend the discussion in (28) a bit further. There will always be a first head, and there will always be a second head, and a third head, and so on. In an experiment where the outcome is always success, with $P(S) = p$, or failure, with $P(F) = q$ ($p + q = 1$), then there will always be a first success, a second success, . . ., a twenty-first success (as in table tennis), and so on. The triangle is involved even here, and the analysis requires looking at the twenty-first diagonal, a formidable undertaking. You might be interested in learning that the most likely number of total points scored in a table tennis match is the minimum of the two numbers

$$\left[1 + \frac{20}{p}\right] \text{ and } \left[1 + \frac{20}{q}\right]$$

where $[x]$ denotes "the greatest integer contained in the number x." Draw the graph of the most likely number of total points as a function of p, for values of p in the interval $0 < p < 1$, and see how sensitive the total is to the skill of the better player.

Open Question

30. We have only scratched the surface of the iceberg. As you progress through your mathematical studies, keep a sharp eye peeled for the most unexpected appearances of the triangle. Sooner or later, you will find another application; the thrill of discovery will be yours, and the best talk of all will be your telling of your discovery. Happy hunting!

REFERENCES AND FURTHER READING

Bryant, J. S., G. E. Graham, and K. G. Wiley. *Nonroutine Problems.* New York: McGraw-Hill Book Co., 1965.

Feller, William. *An Introduction to Probability Theory and Its Applications,* vol. 1. 3d ed. New York: John Wiley & Sons, 1968.

Hoggatt, V. E. *Fibonacci and Lucas Numbers.* Boston: Houghton Mifflin Co., 1969.

Mosteller, F., R. E. Rourke, and G. B. Thomas. *Probability: A First Course.* 2d ed. Reading, Mass.: Addison-Wesley Publishing Co., 1970.

Neff, J. D. "More on Pascal's Triangle." *New York State Mathematics Teachers' Journal* 21 (October 1971): 150–53.

Polya, G. *Mathematical Discovery.* 2 vols. New York: John Wiley & Sons, 1962, 1965.

———. *Mathematics and Plausible Reasoning.* 2 vols. Princeton: Princeton University Press, 1954.

Riordan, John. *An Introduction to Combinatorial Analysis.* New York: John Wiley & Sons, 1958.

Scientific American, especially December 1966, pp. 128–31 and March 1969, pp. 116–20.

Yaglom, A. M., and I. M. Yaglom. *Challenging Mathematical Problems.* 2 vols. San Francisco: Holden-Day, 1964, 1967.

Youse, B. K. *An Introduction to Mathematics.* Boston: Allyn & Bacon, 1970.

6

Topology

Bruce E. Meserve
Dorothy T. Meserve

Topology is a study of sets of points and is often thought of as a basic geometry. In this geometry there are formal rules under which a circle may be transformed into a square, the surface of a teacup into the surface of a doughnut, and even a line segment an inch long into a line segment a mile long. (An inch is as good as a mile in this geometry.)

Many aspects of topology may be observed from such games with geometric figures. Topology may also be studied very seriously and formally. The following topics have been selected to help you gain an informal understanding of some aspects of topology. You can have fun exploring topological ideas without going seriously astray relative to the formalities that are left for possible future study. As you read, be sure to make sketches of figures whenever they can help you visualize the statements under consideration.

Topologically Equivalent Figures

Any two congruent figures are equivalent under a rigid motion, that is, either figure may be mapped onto the other by a transformation that preserves lengths of line segments. If two figures are similar, either figure may be mapped onto the other by a transformation that preserves measures of angles. When lengths of line segments are preserved, each line segment is mapped onto a line segment of the same length as the original line segment. When measures of angles are preserved, each angle is mapped onto an angle of the same measure as the original angle.

In topology we are concerned with transformations that preserve neighborhoods of points. It is possible to give a formal definition of neighborhood, but we shall not attempt to do so. Rather, we shall depend upon your intuitive concept of the word. The neighborhood of a point P on a line may be taken as

an open line segment containing the point P. The neighborhood of a point Q on a plane may be taken as the interior of a circle with Q as center. The neighborhood of a point R in space may be taken as the interior of a sphere with R as center.

If a topological transformation maps a point P onto a point P', then every neighborhood of P (no matter how small) must be mapped into a neighborhood of P' and the points of every neighborhood of P' must be image points from some neighborhood of P. This restriction requires that distinct points correspond to distinct points. Each point of the original figure is mapped onto exactly one point of the new figure, and each point of the new figure is the image of exactly one point of the original figure. Two figures are *topologically equivalent* if neighborhoods are preserved under a mapping of either figure onto the other.

Consider a circle O of radius 3 inches and a circle O' of radius 4 inches. The 3-inch circle is congruent to another 3-inch circle that is concentric with the 4-inch circle. Can you find a pattern for mapping points of the larger of these two concentric circles onto points of the smaller circle so that neighborhoods are preserved? Can you find a pattern for mapping points of the smaller circle onto points of the larger circle so that neighborhoods are preserved? If you have difficulty visualizing a mapping of one of the concentric circles onto the other, make the points along radii correspond, that is, *project* one circle onto the other from their common center. Figure 6.1 illustrates the sort of drawings that you should make as you read this paragraph.

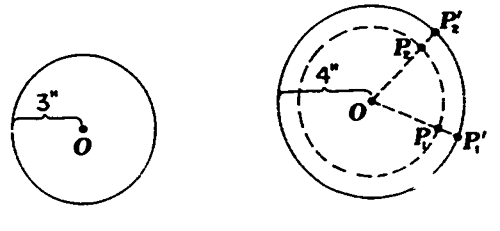

Fig. 6.1

Do you think that any two congruent circles are topologically equivalent? Explain why this must be so. Now you should be able to explain why any circle is topologically equivalent to any other circle. If the words "topologically equivalent" seem too cumbersome, you can say the same thing by saying that the figures are *homeomorphic*.

A *simple closed curve* may be defined as a figure that is topologically equivalent to a circle. Show that any triangle is a simple closed curve; any square, any parallelogram, any regular polygon. Can you think of a curve that is not a simple closed curve? Try a "figure eight," a line segment, a parabola, a hyperbola in our usual geometry, a sine curve. None of these curves are simple closed curves.

Think of a circle tangent to a line at a point A as in figure 6.2, and let B be the point diametrically opposite A. Use lines through B to map (project) every point P different from B on the circle onto a point P' of the line. Neighbor-

hoods are preserved. After the point B has been removed, the remaining *punctured circle* (a circle with a point removed) is topologically equivalent to the line.

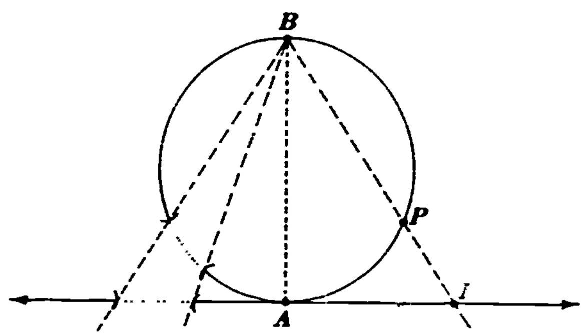

Fig. 6.2

Each of the following statements may be demonstrated as indicated in the figure accompanying the statement.

1. Any open line segment (such as the set of points between 0 and 1 on a number line) is topologically equivalent to a half-line. (In fig. 6.3 where $\overleftrightarrow{DB} \parallel \overleftrightarrow{AC}$, D is a fixed point and P is any point of $\overset{\vdash\!\dashv}{AB}$, observe that $\overset{\vdash\!\dashv}{AB}$ is topologically equivalent to $\overset{\vdash\!\rightarrow}{AC}$.)

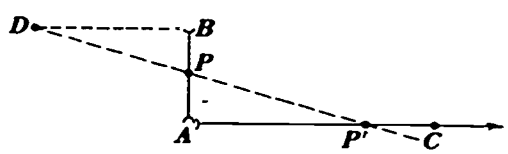

Fig. 6.3

2. Any open line segment is topologically equivalent to a line. (In fig. 6.4 where $\overleftrightarrow{DB} \parallel \overleftrightarrow{AC} \parallel \overleftrightarrow{EF}$, observe that $\overset{\vdash\!\dashv}{BE}$ is topologically equivalent to $\overleftrightarrow{AC}$.)

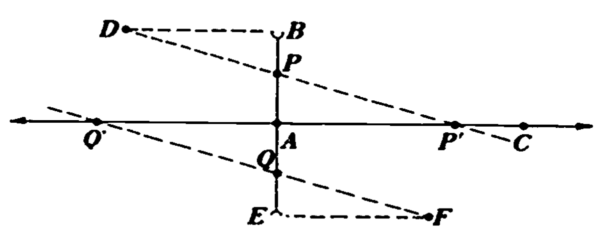

Fig. 6.4

3. Any line segment $\overline{AB}$ is topologically equivalent to any other line segment $\overline{CD}$. (See fig. 6.5.)

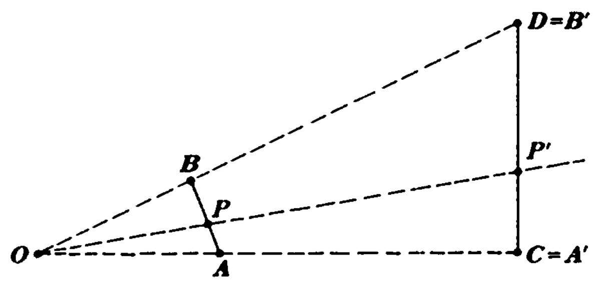

Fig. 6.5

The third statement provides the basis for the comment in the first paragraph of this article that "an inch is as good as a mile in this geometry."

In demonstrations such as those just considered we have tacitly assumed that a figure may be replaced by any congruent figure to obtain suitable figures in an appropriate relative position. This assumption is a special case of the statement that any two figures that are topologically equivalent to the same figure are topologically equivalent to each other. For example, we have already noticed that any 3-inch circle and a 4-inch circle are topologically equivalent to a 3-inch circle concentric with the 4-inch circle and also are topologically equivalent to each other. Draw figures for a few other such examples. Observe the mappings (topol ical equivalences) from the first figure to the second and from the third figure to the second. Is there also a topological equivalence represented by a mapping from the first figure to the third? Explain why the first and third figures must be topologically equivalent. Note that this is a *transitive property* for topological equivalences.

Any figure that is topologically equivalent to a punctured circle is a *simple open curve*. Thus a line is a simple open curve. Explain or show why each of the following statements must be true:

1. An open line segment is a simple open curve.
2. A half-line is a simple open curve.
3. A plane angle is a simple open curve.
4. A parabola is a simple open curve.
5. A sine curve is a simple open curve.
6. An ellipse is not a simple open curve.
7. A hyperbola is not a simple open curve.

The letters of our alphabet may be grouped according to the topological properties of a given representation of the capitals of the letters considered as curves. For example, in their usual style Y and T are topologically equivalent curves. Also M, N, and Z are topologically equivalent curves (see Jacobs, pp. 451–52).

The neat little ink spots known as letters of the alphabet may also be considered as regions on the plane. Then, even though their boundaries are of different shapes, the regions Y, T, M, N, Z, and others are all topologically equivalent to each other and to a circular disk. Similarly, the regions A, O, P, and others are equivalent to each other and to a circular ring; that is, a disk

with a hole in it. In general, any connected bounded plane region with its boundary is topologically equivalent to a disk with some number b of holes. The number b is called the *Betti number* of the region. Any two topologically equivalent regions have the same Betti number.

In space a sphere may be considered as the surface of a solid ball. A sphere is topologically equivalent to the surface of a cube, a tetrahedron, a convex polyhedron, and many other space figures.

Consider the surface of an ordinary doughnut (a *torus*). Because of the hole in the middle, the surface of the doughnut is not equivalent to a sphere. However, the spherical part of a sphere with one handle may be thought of as shrunk into the continuation of the handle. Thus the surface of a doughnut is topologically equivalent to a sphere with one handle. See figure 6.6.

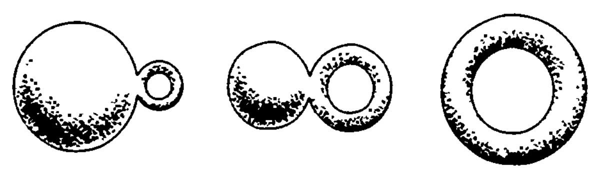

Fig. 6.6

Think of a sphere as a hollow ball such that one hemisphere can be pushed into the other to form a bowl or cup without handles. A sphere without handles is topologically equivalent to a cup without handles. See figure 6.7.

Fig. 6.7

A sphere with one handle is topologically equivalent to a cup with one handle, such as a teacup. Thus an ordinary doughnut is topologically equivalent to a teacup. A sphere with two handles is topologically equivalent to a cup with two handles.

Exercises and further information regarding topological surfaces may be found in several of the references in the bibliography at the end of this chapter.

Traversable Networks

Another famous topological problem is concerned with bridges in the city of Königsberg. There was a river flowing through the city. In the river there were two islands connected to the mainland and each other by seven bridges as shown in figure 6.8.

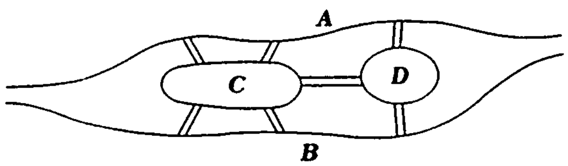

Fig. 6.8

The people of Königsberg loved a Sunday stroll and thought it would be nice to take a walk in which they would cross each bridge exactly once. But no matter where they started or what route they tried, they could not cross each bridge exactly once. This caused considerable discussion. Gradually it was observed that the basic problem was concerned with paths between the two sides A, B of the river and the two islands C, D as in figure 6.9.

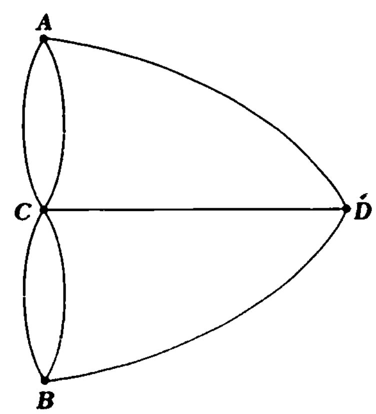

Fig. 6.9

With this geometric representation of the problem it was no longer necessary to discuss the problem in terms of walking across the bridges. Instead one could discuss whether or not the curve associated with the problem is *traversable* in a single trip, that is, whether or not one could start at some point of the curve and traverse each arc exactly once. The curve is often called the *graph* of the problem. The Königsberg bridge problem could be considered in terms of its graph by people who had never even been to Königsberg. The desired walk was possible if and only if the graph was traversable.

When is a graph traversable in a single trip? One can walk around a city block, and it is not necessary to start at any particular point. In general, one may traverse any simple closed curve in a single trip. This may be surprising for some complicated-appearing simple closed curves, but it is a basic property of all simple closed curves.

We next consider walking around two city blocks and down the street separating them. (See fig. 6.10.) This problem is a bit more interesting in that it is necessary to start at B or E. Furthermore, if one starts at B, then one ends at E; if one starts at E, then one ends at B. Note that it is permissible to pass through a vertex several times but each arc must be traversed exactly

once. The peculiar property of the vertices B and E is that each is an endpoint of three arcs while each of the other vertices (A, C, D, F) is an endpoint of exactly two arcs. A similar observation led a famous mathematician by the name of Leonard Euler to devise a complete theory for traversable graphs, often called traversable *networks*.

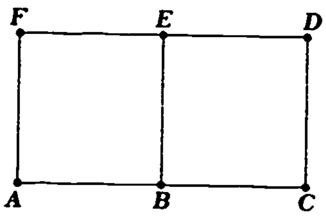

Fig. 6.10

Euler classified the vertices of a graph as odd or even. A vertex that is on (an endpoint of) an odd number of arcs is called an *odd vertex*; a vertex that is on an even number of arcs is called an *even vertex*. Since every arc has two ends, there must be an even number of odd vertices in any graph. Any graph or network that has only even vertices is traversable, and the trip may be started at any vertex. Furthermore, the trip will terminate at its starting point. If a graph contains two odd vertices, the graph is traversable, but the trip must start at one of the odd vertices. The trip will then terminate at the other odd vertex. If a graph has more than two odd vertices, the graph is not traversable in a single trip. In general, a graph with $2k$ odd vertices, where k is a positive integer, may be traversed in k distinct trips.

The graph for the Königsberg bridge problem has four odd vertices. This graph cannot be traversed in a single trip. Thus it was not possible to cross each of the seven Königsberg bridges exactly once in a single trip. The solution of the Königsberg bridge problem is the determination that the desired walk is impossible. The discussion of this problem in an article by Tucker and Bailey (1950) is concluded with the statement that Tucker had actually walked across each of the bridges exactly once in 1935. (There were eight bridges at that time.)

Frequently we see in advanced mathematical theories only complicated manipulations and intricate statements involving precisely worded definitions and theorems. It is refreshing as well as enlightening to look back occasionally at the roots of the theory and see the problems that started great minds working for generalizations that have led to present theories. The Königsberg bridge problem is independent of the size and shape of the objects under consideration. It is a topological problem. It has been considered by some writers to be the starting point of the theory of topology.

The study of networks (graphs) is a part of the major branch of topology that is now called *graph theory* (see Ore). Recent considerations of networks include separate considerations of the point of view of a highway inspector who wishes to traverse each arc (highway) exactly once and the point of view of a salesman who wishes to visit each vertex (town) exactly

once (see Stein). Exercises and further information regarding networks and graph theory may be found in the references in the bibliography.[1]

The Möbius Strip

Let us conclude our consideration of topology with a few comments about a surface that has several very unusual properties. The surface is one-sided. A fly can walk from any point on it to any other point without crossing an edge. Unlike a table top or a wall, it does not have a top and a bottom or a front and a back. This surface is called a Möbius strip and may be constructed very easily from a rectangular piece of paper such as a strip of gummed tape. Size is theoretically unimportant, but a strip an inch wide and about a foot long is easy to handle.

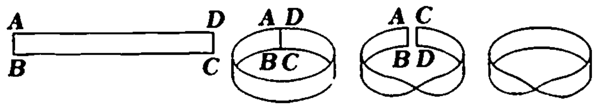

Fig. 6.11

Consider the rectangular strip $ABCD$ shown in figure 6.11. If we simply form a cylinder as *in the second drawing*, then the corners A and D are placed together and the corners B and C are placed together. To construct a Mobius strip we twist the strip of gummed tape just enough to stick the gummed edge of one end to the gummed edge of the other end. In this way we place the corners A and C together and the corners B and D together.

If we cut across the Möbius strip, we again get a single rectangular strip similar to the one we started with. However, if we make a Möbius strip from a rectangular strip and cut around the strip halfway between the long sides of the rectangle (see the dotted line in fig. 6.12), we do not get two strips. Rather we get one strip with two twists in it.

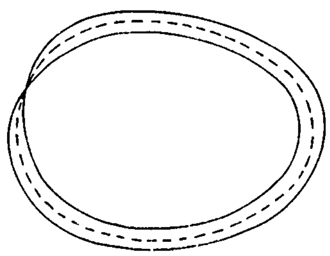

Fig. 6.12

William Hazlett Upson used the peculiar property of Möbius strips in his story called "Paul Bunyan and the Conveyer Belt" (see Fadiman 1962, pp. 33–35). Other stories based on topological concepts may be found (see Fadiman 1958 and 1962).

1. The material in this section is adapted from Meserve 1953, pp. 172–73.

People of all ages can enjoy exploring the properties of a Möbius strip. Make a strip, cut it down the middle to check the property that we have just discussed. Cut the strip down the middle again and see what happens. (You may be astonished.) Make another Möbius strip and try to cut it in thirds; try to cut it in quarters by taking one fourth off the edge. Make a Möbius strip using coat zippers so that it can be cut (unzipped) down the middle once and then down the middle again. One enterprising teacher designed a one-sided dress and a child's bib based on a Möbius strip (see Pedersen). You may find exercises and further discussion of Möbius strips in some of the other references in the bibliography.[1]

Bibliography

Several other topics could have been selected to provide an introduction to topology. Some of you might have enjoyed topics such as the Jordan curve theorem, surfaces and connectivity, Euler's formula, and the four-color problem even more than the topics selected. These and other topological topics may be explored in articles in the *Mathematics Teacher*, a wide variety of books, and, in particular, the following references.

REFERENCES

Arnold, B. H. *Intuitive Concepts in Elementary Topology*. Englewood Cliffs, N. J.: Prentice-Hall, 1962.

Bing, R. H. "Point Set Topology." In *Insights into Modern Mathematics*. Twenty-third Yearbook of the National Council of Teachers of Mathematics, pp. 306-35. Washington, D. C.: The Council, 1957.

Blank, Albert A. "Nets." In *Enrichment Mathematics for High School*. Twenty-eighth Yearbook of the National Council of Teachers of Mathematics, pp. 76-99. Washington, D. C.: The Council, 1963.

Chinn, W. G., and N. E. Steenrod. *First Concepts of Topology*. New Mathematical Library, vol. 18. New York: Random House, 1966.

Courant, Richard, and Herbert Robbins. *What Is Mathematics?* pp. 235-71 New York: Oxford University Press, 1941.

Euler, Leonard. "The Seven Bridges of Koenigsberg." Reprinted in *The World of Mathematics*, vol. 1, pp. 573-80. Edited by James R. Newman. New York: Simon & Schuster, 1956.

Fadiman, Clifton, ed. *Fantasia Mathematica*, pp. 99-109, 155-70, 196 210, 222-36. New York: Simon & Schuster, 1958.

———. *The Mathematical Magpie*, pp. 33-35, 48-59, 70-96, 219-22, 245-50, 263, 292, 293. New York: Simon & Schuster, 1962.

Frechet, Maurice, and Ky Fan. *Initiation to Combinatorial Topology*. Translated by Howard W. Eves. Boston: Prindle, Weber & Schmidt, 1967.

Fujii, John N. *Puzzles and Graphs*. Washington, D.C.: National Council of Teachers of Mathematics, 1966.

Jacobs, Harold R. *Mathematics—a Human Endeavor*, pp. 447-91. San Francisco: W. H Freeman & Co., 1970.

Johnson, Donovan A., and William H. Glenn. *Topology—the Rubber Sheet Geometry*. Manchester, Mo.: Webster Publishing Co., 1960.

1. The material in this section is adapted from Meserve 1953, pp. 473-74.

Kasner, Edward, and James Newman. *Mathematics and the Imagination.* New York: Simon & Schuster, 1940.

Meservé, Bruce E. *Fundamental Concepts of Geometry*, pp. 288–310. Reading, Mass.: Addison-Wesley Publishing Co., 1955.

———. "Topology for Secondary Schools." *Mathematics Teacher* 46 (November 1953): 465–74.

Meserve, Bruce E., and Max A. Sobel. *Mathematics for Secondary School Teachers*, pp. 338–56. Englewood Cliffs, N. J.: Prentice-Hall, 1962.

Ore, Oystein. *Graphs and Their Uses*. New Mathematical Library, vol. 10. New York: Random House, 1963.

Pedersen, Jean J. "Dressing Up Mathematics." *Mathematics Teacher* 61 (February 1968). 118–22.

Stein, Sherman K. *Mathematics: The Man-made Universe*, pp. 125–39, 212–37. 2d ed. San Francisco: W. H. Freeman & Co., 1969.

Steinhaus, Hugo. *Mathematical Snapshots*, pp. 252–81. 3d American ed. New York: Oxford University Press, 1969.

Tietze, Heinrich. *Famous Problems of Mathematics*, pp. 64–89, 226–42. Translated from 2d German ed. by Beatrice Kevitt Hofstadter and Horace Komm. Baltimore: Graylock Press, 1965.

Tucker, A. W., and H. S. Bailey, Jr. "Topology." *Scientific American* 182 (January 1950): 18–24.

7

Experiments with Natural Numbers

Richard V. Andree

The natural numbers $1, 2, 3, 4, \ldots, N, \ldots$ hold their secrets well. The ancient Greeks studied them with considerable enterprise. Euclid, of geometry fame, actually published more work on number theory than he did on geometry; and many of his proofs still stand today as models of ingenuity.

Problems involving the positive integers have long interested educated amateurs as well as professional mathematicians. Several important discoveries have been made by amateurs and even by schoolboys. No doubt there are many more still to be discovered. If your school has a desk calculator or a computer, or even a set of tables, you, too, may participate in the thrill of numerical exploration and discovery. Someone else may have made the same discoveries earlier, but that does not take away from either the thrill of discovery or the credit due you—providing *you* make the discoveries and create your own proofs rather than looking them up in the library.

Are you ready? Let's go.

Excursion 1—Sums

The sums of the odd integers (those not evenly divisible by 2, namely, 1, 3, 5, 7, 9, . . .) show an interesting property:

$$
\begin{aligned}
&1 &&= 1^2.\\
&1 + 3 &&= 4 = 2^2.\\
&1 + 3 + 5 &&= 9 = 3^2.\\
&1 + 3 + 5 + 7 &&= 16 = 4^2.\\
&1 + 3 + 5 + 7 + 9 &&= 25 = 5^2.\\
&1 + 3 + 5 + 7 + 9 + 11 &&= 36 = 6^2.\\
&1 + 3 + 5 + 7 + 9 + 11 + 13 &&= 49 = 7^2.\\
&1 + 3 + 5 + 7 + 9 + 11 + 13 + 15 &&= 64 = 8^2.
\end{aligned}
$$

From this, one might jump to the following conjecture:

The sum of the first K odd integers is K^2.

This conjecture may or may not be correct. Your problem is to either prove or disprove the conjecture. If a computer or desk calculator (even an adding machine) is available, we can easily check our conjecture for the first few values. The following program, written in the BASIC computer language, will check our conjecture for the sum of the first K = 1, 2, 3, . . ., 50 odd integers.

```
10    PRINT "K", "K*K", "SUM OF FIRST K ODD INTEGERS
20    S = 1
30    N = 1
40    FOR K = 2 TO 50
50    N = N + 2
60    S = S + N
70    PRINT K, K*K, S
80    NEXT K
90    PRINT "END OF PROGRAM"
100   END
:
```

A partial output of this program is shown below:

```
:RUN
K              K*K            SUM OF FIRST K ODD INTEGERS
 2              4              4
 3              9              9
 4             16             16
 5             25             25
 6             36             36
 7             49             49
 8             64             64
 9             81             81
10            100            100
11            121            121
12            144            144
13            169            169
14            196            196
15            225            225
16            256            256
17            289            289
18            324            324
19            361            361
20            400            400
                .              .
 .              .              .
 .              .              .
48           2304           2304
49           2401           2401
50           2500           2500
END OF PROGRAM
END PROGRAM
:
```

Since typing of results is rather slow and wastes computing time, the following program (which types only when S ≠ K*K) will test the conjecture for K = 1, 2, . . ., 1000 in about the same time that the first program required to print the first ten lines of output.

```
10  PRINT" THE ONLY VALUES OF K <= 1000 FOR WHICH THE SUM OF
      THE"
11  PRINT "FIRST K ODD INTEGERS DOES NOT EQUAL K*K ARE LISTED
      BELOW:"
30  S = 1
40  N = 1
50  FOR K = 2 TO 1000
55  N = N + 2
60  S = S + N
70  IF S = K*K THEN 90
80  PRINT K, K*K, S
90  NEXT K
95  PRINT "       END OF JOB"
110 END
```

The output of the program is:

```
 RUN
 THE ONLY VALUES OF K <= 1000 FOR WHICH THE SUM OF THE
FIRST K ODD INTEGERS DOES NOT EQUAL K*K ARE LISTED BELOW:
        END OF JOB
END PROGRAM
```

Thus, we have shown by actual experiment that the conjecture "The sum of the first K odd integers is K^2" is valid for all $K \leq 1000$, but that still does not *prove* it is a true statement for all positive integers K, even if it does bolster our confidence.

The actual proof is not beyond your ability. It can be done easily by mathematical induction or by using an ingenious geometric argument based on reasoning similar to that suggested in the diagram shown in figure 7.1.

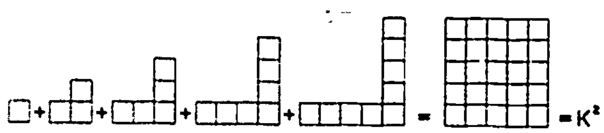

Fig. 7.1

On this first excursion we have guided you rather carefully over the ground to be traversed and planted clear guideposts along the way. The final proof is still left for you. Some of our later excursions will prove more adventurous. Some will even lead you into uncharted waters. We hope you will participate actively in the adventure and let your own imagination wander a bit, too. For example, what, if anything, can you discover about the sum of the first K even integers? (Can you *prove* your conjecture?) What about the sum of the first K integers (both odd and even)? You are probably already aware that the answer is $K(K + 1)/2$, but have you ever *proved* it?

Even more ingenuity is needed to prove the following conjectures:

The sum of the squares of the first K positive integers is $K(K + 1)(2K + 1)/6$.

The sum of the cubes of the first K positive integers is the square of the sum of the first K positive integers, i.e., $1^3 + 2^3 + 3^3 + \ldots + K^3 = (1 + 2 + 3 + \ldots + K)^2$.

Can you make (and prove??) similar conjectures about the sums of the squares (or cubes) of the first K odd (even) integers, or about higher powers of the integers? Do a bit of investigating on your own. A set of tables or a

desk calculator or a small computer would be helpful but is not really essential.

Excursion 2—Palindromes

Positive integers such as 4735374 or 461164 that read the same forward as backward are called palindromes. An interesting number game is described below:

1. Take a positive integer N.
2. Form the sum of N and its reverse. (The reverse of N is the number obtained writing the digits of N in reverse order. The reverse of 826 is 628.)
3. If the sum is a palindrome, STOP, otherwise continue with step 2, using the sum as the new value of N.

Example 1:

1. N = 139.
2. New N = 139 + 931 = 1070.
3. 1070 is not a palindrome.
2. 1070 + 0701 = 1771.
3. 1771 is a palindrome in two cycles.

Example 2:

N = 48017

48017 + 71084 = 119101, not a palindrome.
119101 + 101911 = 221012, not a palindrome.
221012 + 210122 = 431134, a palindrome in three cycles.

However, forming a palindrome is not always so easy; 89 does not produce a palindrome until 24 cycles have been completed, and 196 goes a long, long time without producing a palindrome. I don't know if it ever does. Investigate whether or not N = 5, 6, . . ., 99 all produce palindromes and if so, how long the cycles are. If you wish, extend your investigations to starting values of more than two digits. Can you determine several infinite sets of starting values that will always produce palindromes in one cycle? In two cycles? What else can you discover about starting values that produce palindromes? As far as I know, no one has ever proved that every starting value will eventually produce a palindrome, but neither has anyone produced a starting value that they can guarantee does not eventually produce a palindrome. You may wish to consider how such a guarantee could be given.

Excursion 3—Related Digits

We now turn our attention to a number of problems that are related to the individual digits that make up certain positive integers. Our first problem is as follows:

Determine all of the three-digit integers for which the sum of the cubes of the digits of the number equals the number. Since

$$1^3 + 5^3 + 3^3 = 1 + 125 + 27 = 153,$$

we know such numbers do exist.

It is possible to solve this problem using pencil and paper. However, if a computer or calculator is available, we let H, T, and U represent the hundreds, tens, and units digits. The original number is N = 100*H + 10*T + U, and the sum of the cubes of the digits is S = H ↑ 3 + T ↑ 3 + U ↑ 3 in BASIC notation.

The following program suggests one possible "brute force" technique, in which we simply try all possible three-digit numbers to see if N = 100*H + 10*T + U = H ↑ 3 + T ↑ 3 + U ↑ 3 and print out those N that do satisfy the given condition. We let H take on the values from 1 to 9 while T and U take on values from 0 to 9. (Why?) The values HTU advance much as a car odometer would in going from 100 to 999, with tests occurring at each value.

Study the program given below before continuing.

```
10   REM FIRST TRIAL AT SUM OF CUBES OF DIGITS = NUMBER
20   FOR H = 1 TO 9
30     FOR T = 0 TO 9
40       FOR U = 0 TO 9
50         LET N = 100*H + 10*T + U
60         LET S = H ↑ 3 + T ↑ 3 + U ↑ 3
70         IF S<>N THEN 80
75         PRINT N;
80       NEXT U
90     NEXT T
100  NEXT H
110  PRINT, "     END OF EXAMPLE"
150  END
:
```

This program does not use statement 75 to print out a successful find until H = 1, T = 5, U = 3. At that time S = N since $153 = 1^3 + 5^3 + 3^3$. Therefore, in statement 70 the transfer to statement 80 is not made. Line 75 is then executed. Your program should type out

```
153     370     371     407     END OF EXAMPLE
```

The actual computer time used to do the problem (approximately 10.5 seconds on the Nova) will vary from computer to computer, but the vital thing is to notice how much faster it could run using a program that is no harder to write but uses "common sense" in the way it attacks the problem.

The above program ran and produced correct answers; but as a computer program, it is at best inept. By using a very small amount of "programming common sense," you could have saved more than 50 percent of the computation time on this problem. Are you ingenious enough to see it before you continue?

Write out your revised program before continuing.

The program computes

$$N = H^3 + T^3 + U^3$$

each time it goes through the inside FOR U = 0 TO 9 . . . NEXT U-loop for a total of $3 \times 2 \times 900 = 5400$ time-consuming multiplications and 1800 additions. If each digit were cubed directly after the FOR loop for that digit, it would cut this down to $18 + 180 + 1800 = 1998$ multiplications and the same 1800 additions (which take much less time than multiplications do).

This saving of 3402 multiplications shows the difference between a complete "hack" programmer and a capable one.

We can also save some time by using H*H*H rather than the slower H ↑ 3 instruction. On computers that compute H ↑ 3 as EXP (3*LOG(H)), increased accuracy may also result. The following program will do this.

```
10   REM MORE POLISHED VERSION OF SUM OF CUBES OF DIGITS = NUMBER
20   PRINT "THREE DIGIT INTEGERS N HAVING THE SUM OF THE CUBES OF THE"
30   PRINT "DIGITS OF N EQUAL TO N ARE;"
50   FOR H = 1 TO 9
60     LET H3 = H*H*H
70       LET N1 = 100*H
80       FOR T = 0 TO 9
90         LET T3 = T*T*T
100          LET N2 = 10*T
110          FOR U = 0 TO 9
120            LET S = H3 + T3 + U*U*U
130          LET N = N1 + N2 + U
140              IF S<>N THEN 150
145  PRINT N;
150            NEXT U
160          NEXT T
170        NEXT H
180  PRINT
190  PRINT "END OF EXAMPLE"
230  END
```

The output is

```
RUN
THREE DIGIT INTEGERS N HAVING THE SUM OF THE CUBES OF THE
DIGITS OF N EQUAL TO N ARE;
 153      370      371      407
END OF EXAMPLE
END PROGRAM
```

The time used in this revised version is 4.7 seconds, as compared with 10.5 seconds for the original program. This may seem a small savings, but it is a savings of over 50 percent—and that is very worthwhile on long or frequently run programs. Students should think about *efficient* use of the computer if they plan to use it. The use of human intelligence as a computer-saving device is what makes one programmer worth three times as much as another with the same experience. If you can save only ten minutes per working day on a computer worth $600 an hour, you have *saved* between $2000 and $3000 per month. No wonder employers are willing to pay an extra $1000 per month to a really able programmer over and above what they will pay an ordinary programmer! Which will you be? Try the following:

1. Our programs determine all the *three-digit* integers such that the sum of the cubes of those digits is equal to the original number. However, there may be one- and two-digit integers that also have this property. Modify the above program so that it will determine all one-, two-, and three-digit integers having the above property.

2. Use some mathematical ingenuity to supplement your computer output

and determine *all* integers (no matter how many digits) that have the property that the sum of the cubes of the digits of the number equals the original number.

3. Determine all the positive integers N such that the *sum of the factorials* of the digits of N equals the original number N. Be clever in your programming. Perhaps you may even wish to use subscripted variables and compute and store the values of 0!, 1!, 2!, 3!, . . . 9! in subscripted storage locations to reduce the amount of computation.

Here is another little-known problem concerning digits that merits investigation:

Let N_0 be any four-digit positive integer. If we define the following relations as

$$L_K = \text{(the largest integer obtainable by rearranging the digits of } N_K\text{)}$$

and

$$S_K = \text{(the smallest integer obtainable by rearranging the digits of } N_K\text{)}$$

then $N_{K+1} = L_K - S_K$.

Thus if

$$\begin{aligned} N_0 &= 7162, \\ N_1 &= 7621 - 1267 = 6354, \\ N_2 &= 6543 - 3456 = 3087, \\ N_3 &= 8730 - 0378 = 8352, \\ N_4 &= 8532 - 2358 = 6174, \end{aligned}$$

and

$$N_5 = 7641 - 1467 = 6174,$$

which clearly repeats forever.

Also, if

$$\begin{aligned} N_0 &= 2212, \\ N_1 &= 2221 - 1222 = 0999, \\ N_2 &= 9990 - 0999 = 8991, \\ N_3 &= 9981 - 1899 = 8082, \\ N_4 &= 8820 - 0288 = 8532, \\ N_5 &= 8532 - 2358 = 6174, \\ N_6 &= 7641 - 1467 = 6174, \end{aligned}$$

which again clearly repeats forever.

In contrast, if

$$\begin{aligned} N_0 &= 7777, \\ N_1 &= 7777 - 7777 = 0, \end{aligned}$$

which also repeats forever.

Your problem is to investigate this recurrence relation. The results may surprise you.

It is *not* necessary to investigate all 9000 possible 4-digit numbers to completely solve this problem. The 24 starting values 7162, 7621, 1726, 2671, etc., each yield the same value for N_1 and thus the same sequence from there on. Show that there are only 54 possible values for N_1, and therefore we need spend less than 1/100 of the time that would be required to check all 9000 cases. You

may also wish to investigate this process for starting values having 3 or 5 or 6 digits. Some 6-digit starting values lead to the repeater 631764, others do not. Note that $165033 = 16^3 + 50^3 + 33^3$, base 100 instead of base 10.

Excursion 4—Factorials

If N is a positive integer, then the product of the positive integers from 1 to N inclusive is called *N-factorial* and symbolized as

$$N! = 1 \cdot 2 \cdot 3 \cdot 4 \cdot \ldots \cdot (N - 1) \cdot N$$

In some problems it is also desirable to define 0! as 1, but we shall not.

Here are some problems for you to investigate.

1. It is true (but *not* obvious) that all ten possible digits appear among the digits of N! for some integers N. The table seen in figure 7.2 lists the smallest positive integer N_K that contains the digit K for K = 0, 1, 2, 3, 4, 5, 6, 7, 8, 9. First verify that the table as given is correct (or correct it) and then extend it to larger values of K. Note, for example, that when K = 12, N = 5 since 5! = 120. Can you extend the table as far as K = 100?

K	0	1	2	3	4	5	6	7	8	9
Smallest N such that N! contains K	5	1	2	8	4	7	3	6	9	11
N!	120	1	2	40320	24	5040	6	720	362880	39916800

Fig. 7.2

2. What is the smallest positive integer N that contains all 10 of the digits (in some order) among the digits of N!?

3. Are there values of N other than N = 1, 2 such that N! begins with N? Note that the computer may be able to produce a "Yes" answer, but cannot give a definite "No" answer. Why not?

4. J. Maxfield has shown (*Math Magazine*, Vol. 43, #2, March 1970, pp. 64–67) that given a sequence of digits $K = d_1, d_2, d_3, \ldots, d_n$, there exists an integer N such that N! begins with the sequence of digits K. Write a program to create a table such that the *smallest* such positive integer N(K) is determined for each $K \leq 100$. You may wish to also print out the floating point value of N! if it will be available at no additional cost. Eventually, we would like a formula for N(K) = smallest positive integer N such that N! begins with the sequence of digits K, but that may be too much to hope for. Your table might well begin:

K	N(K)	N!
1	1	1
2	2	2
3	9	362880
4	8	40320
5	7	5040
6	3	6
7	6	720
8	14	$8.7178\ldots \times 10^{10}$
9	96	$9.9168\ldots \times 10^{149}$
10	27	.
11	22	.
12	5	.
13	15	
.	.	
.	.	
.	.	
100		

5. If $N \geqq 5$, then $N!$ ends with one or more "trailing zeros." Create a rule that will enable you to look at N and determine *exactly* how many trailing zeros $N!$ will have.

6. The numeral that represents 100! contains exactly 158 digits. Is there a value of N such that $N!$ contains exactly 100 digits?

Excursion 5—Superprimes

We next turn our attention to an interesting subset of the primes called the *superprimes*.

The integer 7331 is a prime. So is any integer obtained by deleting digits from the right edge of 7331, since 733, 73, and 7 are each prime.

A prime integer that has the property that every integer obtained by deleting an arbitrary number of its right-most digits is again prime is called a *superprime*. The integers 317 and 2399 are also superprimes. (You should verify this—it is part of reading mathematics.)

The superprimes 7331 and 317 are rather special, even among superprimes, and are called *superprime leaders* since there is no digit X that will make either 73,11X or 3,17X into a superprime. (If you wish to do some mathematical research get out your pencil and paper and *prove* this last statement before you continue reading. It isn't very difficult.)

A superprime leader is a superprime that cannot be obtained by deleting digits from a larger superprime. The superprime 2399 is not a superprime leader, since 23993 is also a superprime. Actually 23993 is not a superprime leader either.

Ascertain which four of the following seven integers are superprime leaders:

59 23339 59393339 7323 7331 239933 739397.

You may save considerable time and effort if you are observant enough to make and prove the validity of a few observations (that is, theorems) along the way. The types of theorems that could be helpful might include the following:

THEOREM. *If a superprime contains either of the digits 2 or 5, then 2 or 5 can only be the left-most digit of the superprime.*

THEOREM. *The digits 4, 6, 8, 0 will not appear in a superprime.*

THEOREM. *The digit 1 cannot appear as the left-most digit of a superprime.* (This is true because 1 is neither prime nor composite.)

These and other theorems that are not difficult to prove can be most helpful in any search for superprime leaders.

Research proposition 1

Select an integer $K \geqq 4$ and determine all of the superprime leaders having K or fewer digits. If you have a computer available you may wish to use it. If not, use prime tables, which should be available in your library.

For those of you who would like some assurance that you are progressing satisfactorily the following partial table is presented.

Number of digits	2	3	4	5	6	7	8
Number of known superprime leaders having the above number of digits	1	3	4	6	5	3	5

Research proposition 2

Determine whether the number of superprime leaders is finite or infinite. Don't guess. This requires a proof, but the proof is within your ability if you have solved research proposition 1 for sufficiently large K.

Excursion 6—A Recursive Function

The problem given here is one that has not as yet (1972) been investigated very fully. Perhaps you or some of your classmates may be able to add to our current knowledge of this little problem, which arose while creating an example of an integral recursive function for a computer class being taught to high school students on Saturdays on the campus of the University of Oklahoma.

The problem is simple.

Let X_0 be any integer.

We now obtain a sequence of integers starting with X_0 using the recursive rule.

$$X_{K+1} = \begin{cases} X_K/2 & \text{if } X_K \text{ is even} \\ 3X_K + 1 & \text{if } X_K \text{ is odd} \end{cases}$$

If $X_0 = 11$, the corresponding sequence is 11 34 17 52 26 13 40 20 10 5 16 8 4 2 1. If a chain contains the digit 1, it repeats the digits 1 4 2 1 4 2 1 4 2 . . . in a periodic fashion thereafter. In this case we terminate the chain at the first 1 and say the chain converged to 1. Furthermore, we define the number of steps from X_0 to the first 1 to be the length of the chain; that is, if the integer 1 appears in the sequence for the first time as X_α, then the length of the chain is α. The above chain, which begins with $X_0 = 11$, has $\alpha = 14$.

Several typical chains follow.

25 76 38 19 58 29 88 44 22 11 34 17 52 26 13 40 20 10 5 16 8 4 2 1
26 13 40 20 10 5 16 8 4 2 1
27 82 41 124 62 31 94 47 142 71 214 107 322 161 484 242 121 364 182 91 274 137 412 206 103 310 155 466 233 700 350 175 526 263 790 395 1186 593 1780 890 445 1336 668 334 167 502 251 754 377 1132 566 283 850 425 1276 638 319 958 479 1438 719 2158 1079 3238 1619 4858 2429 7288 3644 1822 911 2734 1367 4102 2051 6154 3077 9232 4616 2308 1154 577 1732 866 433 1300 650 325 976 488 244 122 61 184 92 46 23 70 35 106 53 160 80 40 20 10 5 16 8 4 2 1
28 14 7 22 11 34 17 52 26 13 40 20 10 5 16 8 4 2 1
29 88 44 22 11 34 17 52 26 13 40 20 10 5 16 8 4 2 1
30 15 46 23 70 35 106 53 160 80 40 20 10 5 16 8 4 2 1

It will be noted that each of the given chains terminates with the integers 16, 8, 4, 2, 1. This makes us wonder if every chain of length 4 or greater will end with this sequence. We do not know the answer to this question. However, it is easy to prove the following theorem:

THEOREM. *If a chain of the sequence*

$$X_{K+1} = \begin{cases} X_K/2 & \text{if } X_K \text{ is even} \\ 3X_K + 1 & \text{if } X_K \text{ is odd} \end{cases}$$

does converge to 1, then its terminal elements will be 16, 8, 4, 2, 1 if it has length 4 or greater.

The proof is given by working backward from 1, and it is left for the reader. The rub, of course, is that we do not yet know that for every starting value X_0, the chain will converge to 1. Actual experiment (on a computer, of course) shows that all $X_0 < 10{,}000$ do converge to 1, but this does not prove the conjecture that all chains eventually converge to 1.

Another theorem that is important sounding but easy to prove is:

THEOREM. *Given an integer $z > 0$, there exists a starting value X_0 such that the length of the chain from X_0 to 1 is z.*

It is left for the reader to discover a simple formula that yields a starting value $X_0(z)$ that will produce a chain of length z.

Unfortunately no formula that will enable one to look at X_0 and easily determine the length of the chain from X_0 to 1 is known for general X_0. Perhaps you or one of your classmates will eventually solve this problem, but let us put it aside for now and look at some other problems.

If there is a starting value X_0 that does not converge to 1 (and there could well be either a value that "cycles" somewhere else or one that increases without bound, as far as our knowledge given here shows) then there must be a smallest X_0 that does not converge to 1. It is almost obvious that such a "smallest X_0 that does not converge to 1" cannot be even. (Why not?)

Your research problem is to investigate the recursive function

$$X_{K+1} = \begin{cases} X_K/2 & \text{if } X_K \text{ is even} \\ 3X_K + 1 & \text{if } X_K \text{ is odd} \end{cases}$$

and to prove as many theorems about it as possible. As far as your author knows no one has yet (1973) proved that for every integral starting value $X_0 > 0$, the sequence always converges to 1, nor has anyone yet devised a formula to determine the length z of a chain from X_0 to 1 that works for an

arbitrary starting value N_0. Still, there are a number of theorems that can be proved. Thinking up possible theorems (call them conjectures) and then trying to either prove or disprove them is an important phase of mathematical research.

Another recursive function you may enjoy investigating is:

$$\begin{aligned}
N_0 &= 92463\\
N_1 &= 81 + 4 + 16 + 36 + 9 = 146\\
N_2 &= 1 + 16 + 36 = 53\\
N_3 &= 25 + 9 = 34\\
N_4 &= 9 + 16 = 25\\
N_5 &= 4 + 25 = 29\\
N_6 &= 4 + 81 = 85\\
N_7 &= 64 + 25 = 89\\
N_8 &= 64 + 81 = 145\\
N_9 &= \quad = 42\\
N_{10} &= \quad = 20\\
N_{11} &= \quad = 4\\
N_{12} &= \quad = 16\\
N_{13} &= \quad = 37\\
N_{14} &= \quad = 58\\
N_{15} &= \quad = 89, \text{ which repeats in a cycle.}
\end{aligned}$$

Show that for *every* starting value N_0 the

$$N_{K+1} = \text{the sum of the squares of the digits of } N_K$$

recurrence relation either converges to 1 (which then repeats) or converges to the cycle of length 8 that is given above. This is an interesting problem, and the proof requires ingenuity—but no mathematics beyond ninth-grade algebra.

While this particular problem has now been solved, the related problems of N_{K+1} = the sum of the cubes of the digits of N_K and in general

$$N_{K+1} = \text{the sum of the nth powers of the digits of } N_K$$

have not been investigated.

Excursion 7—Zeros

There are integers such as 10000 or 3000000 whose leading digit is (of course) nonzero, but all *the rest of whose digits are zero*. Some such numbers can be factored into two factors, *neither of which contains any zeros at all.*

$$10000 = 16 * 625$$

is such an integer, as is

$$1000000000 = 1953125 * 512.$$

However,

$$100000000 \text{ and } 50000000$$

each contain a zero somewhere *in every possible factorization* into two factors. Your problem is to investigate the phenomenon in as much detail as you can. Start with only those numbers having 8 or fewer digits. It may be well to start your investigation with the additional restriction that the initial (and only nonzero) digit is a 1, that is, that your starting value is a power of 10. As far as your author knows the only powers of $10 \geqq 10^{33}$ that possess a factorization in which neither factor contains a zero are 10^N for

$$N = 1, 2, 3, 4, 5, 6, 7, 9, 18, 33.$$

It is quite possible that there *are* others with exponent $N > 33$. The following seem to be the *only likely candidates* for $N < 10500$ (and quite possibly the only ones for exponent $N < 50{,}000$), but your author has only checked the ones listed above.

$$N \stackrel{?}{=} 34, 35, 36, 37, 39, 49, 51, 67, 72, 76, 77, 81, 86.$$

The related problem for leading digits other than 1 seems not to have been investigated at all. What can *you* discover?

Excursion 8—Irrational Approximations

A rational number is a quotient N/D of two integers with $D > 0$. There are many real numbers such as $\sqrt{2}$ that *cannot equal* a rational number. A proof of this can be found in most high school advanced algebra books.

Even though $\sqrt{2}$ can never equal a rational number N/D, it is quite possible to determine rational numbers N/D that approximate $\sqrt{2}$ (or any other $\sqrt{M}$, $M > 0$) *as closely as is desired.*

Let us find a rational number N/D, with $D > 0$ such that $|(N/D)^2 - 2| < .001$. An algorithm (method) for determining such a rational number N/D is given in the flow chart seen in figure 7.3. The algorithm can be carried out mechanically on a small computer or desk calculator. It can be carried out in many fewer steps if instead of following it mechanically you "play your hunches" on how much to change N and D at each step to approach the desired approximation. Such intuitive involvement becomes important and helpful if the user is working by hand or with a nonprogrammed desk calculator. If a programmable calculator or computer terminal is used, the suggested routine is easy to program and will produce the desired results at about 181/128, which produces 1.41406 as N/D and 1.9996 as $(N/D)^2$. More alert students may wish to discuss the fact that $(N/D)^2$ and N^2/D^2 need not be identical on the computer.

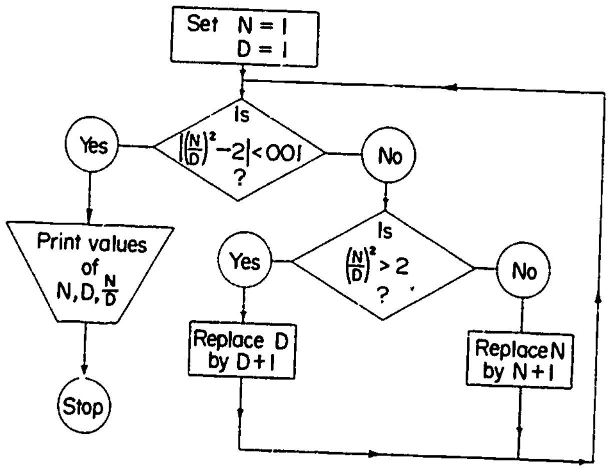

Fig. 7.3

A much faster algorithm for actually computing $\sqrt{M}$, $M > 0$ is given in figure 7.4, showing Newton's method, but the above algorithm seems to produce greater understanding in many students. Both algorithms merit study.

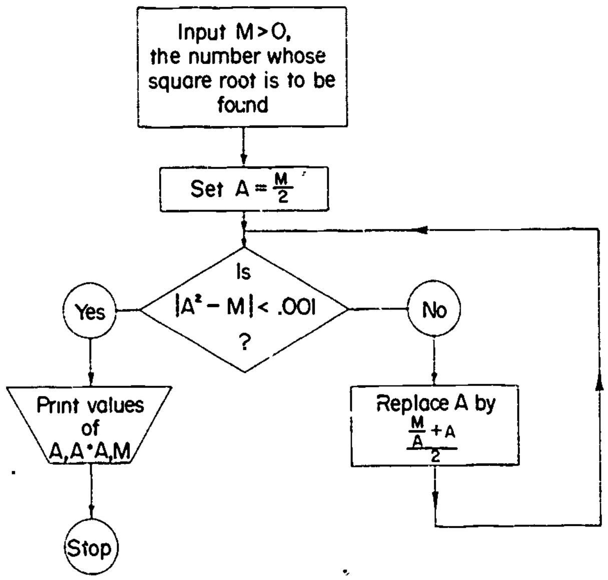

Fig. 7.4

In this case the approximation of $\sqrt{M}$ as A is replaced by the average of A and M/A to obtain a closer approximation. The complete proof that $(A + M/A)/2$ is usually a better approximation of $\sqrt{M}$ than is A is not easy. Most students will agree that it at least seems *likely* that it may be an improvement.

A number of investigations should be carried out on each of the above algorithms. How fast does each obtain an approximation? What happens along the way (insert additional print statements)? What would happen if $M \leq 0$ on algorithm 2? Can you revise algorithm 1 to find $\sqrt{M}$, $M > 0$ rather than $\sqrt{2}$? Can you improve either or both algorithms by permitting the user to insert a "first guess" for N, D (algorithm 1) or A (algorithm 2)? Does this really save any time on a computer terminal or does it take so long to insert the first guess that the computer could obtain several hundred approximations during that time and hence be closer to the answer than your guess?

Excursion 9—Goldbach's Conjectures

Many years ago (1742) Goldbach made two conjectures: (1) Every even integer $N \geq 6$ is the sum of two odd primes. (2) Every odd integer $N \geq 9$ is the sum of three odd primes.

If the first conjecture is true, then the second conjecture can easily be proved. (Can you do it? Try!) However, no one has yet been able to prove the first conjecture. N. Pipping has verified the first conjecture for all $N \leq$ 100,000 by producing examples, but this is of little assistance in proving that

it is valid for all N. In 1937, Vinogradov proved that there exists an integer K such that, for all odd N > K, the second conjecture is valid. However, again, this is an existence proof and no one knows how large K actually is. Recent results show that K is less than $10^{400,000}$, thus it might seem possible to examine all integers less than $10^{400,000}$ on a computer to complete the proof—however, $10^{400,000}$ is a large number, and the task is formidable. Viggo Brun proved that every positive even integer N can be written as the sum of two positive odd integers, each of which is the product of nine or fewer prime factors. Recently, it has been possible to reduce the "nine" in this result to "four," but this is still far short of Goldbach's conjecture.

Actually, there are even integers such as 20 that can be represented as a sum of two odd primes in more than one way:

$$20 = 13 + 7.$$
$$20 = 17 + 3.$$

Our problem on this excursion is to express each integer between 6 and 1,000 as a sum of two odd primes in as many ways as possible. Many books of tables contain lists of primes which will be helpful.

The following BASIC computer program, which expresses each even number between 6 and 100 as a sum of two odd primes, can be extended rather easily, or you can write a more efficient program of your own.

```
10    REM    THERE ARE 24 ODD PRIMES < 100
20    DATA    3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, 53, 59, 61, 67
21    DATA 71, 73, 79, 83, 89, 97
30    DIM P(24)
40    FOR I = 1 TO 24
50    READ P(I)
60    NEXT I
70    FOR N = 6 TO 100 STEP 2
80    FOR I = 1 TO 24
90    LET L = P(I)
100   FOR J = 1 TO I
110   U = P(J)
120   IF L + U < > N THEN 150
130   PRINT N; "="; U; "+"; L,
140   GO TO 170
150   IF L + U > N THEN 170
160   NEXT J
170   NEXT I
180   NEXT N
190   PRINT
200   PRINT "END OF PROBLEM"
250   END
```

The output for this program follows:

6 = 3 + 3	8 = 3 + 5	10 = 5 + 5	10 = 3 + 7
12 = 5 + 7	14 = 7 + 7	14 = 3 + 11	16 = 5 + 11
16 = 3 + 13	18 = 7 + 11	18 = 5 + 13	20 = 7 + 13
20 = 3 + 17	22 = 11 + 11	22 = 5 + 17	22 = 3 + 19
24 = 11 + 13	24 = 7 + 17	24 = 5 + 19	26 = 13 + 13
26 = 7 + 19	26 = 3 + 23	28 = 11 + 17	28 = 5 + 23
30 = 13 + 17	30 = 11 + 19	30 = 7 + 23	32 = 13 + 19
32 = 3 + 29	34 = 17 + 17	34 = 11 + 23	34 = 5 + 29

```
34 =  3 + 31
36 =  5 + 31
40 = 11 + 29
42 = 11 + 31
44 =  3 + 41
46 =  3 + 43
48 =  7 + 41
50 =  7 + 43
52 =  5 + 47
54 = 11 + 43
56 =  3 + 53
58 =  5 + 53
60 = 17 + 43
62 = 19 + 43
64 = 11 + 53
66 = 23 + 43
66 =  5 + 61
70 = 23 + 47
72 = 31 + 41
72 = 11 + 61
74 = 13 + 61
76 = 23 + 53
78 = 37 + 41
78 = 11 + 67
80 = 19 + 61
82 = 29 + 53
84 = 41 + 43
84 = 17 + 67
86 = 43 + 43
86 =  3 + 83
88 =  5 + 83
90 = 29 + 61
90 = 11 + 79
92 = 13 + 79
94 = 23 + 71
96 = 37 + 59
96 = 13 + 83
98 = 19 + 79
100 = 17 + 83
36 = 17 + 19
38 = 19 + 19
40 =  3 + 37
42 =  5 + 37
46 = 23 + 23
48 = 19 + 29
48 =  5 + 43
50 =  3 + 47
54 = 23 + 31
54 =  7 + 47
58 = 29 + 29
60 = 29 + 31
60 = 13 + 47
62 =  3 + 59
64 =  5 + 59
66 = 19 + 47
68 = 31 + 37
70 = 17 + 53
72 = 29 + 43
72 =  5 + 67
74 =  7 + 67
76 = 17 + 59
78 = 31 + 47
78 =  7 + 71
80 = 13 + 67
82 = 23 + 59
84 = 37 + 47
84 = 13 + 71
86 = 19 + 67
88 = 41 + 47
90 = 43 + 47
90 = 23 + 67
90 =  7 + 83
92 =  3 + 89
94 = 11 + 83
96 = 29 + 67
96 =  7 + 89
100 = 47 + 53
100 = 11 + 89
36 = 13 + 23
38 =  7 + 31
42 = 19 + 23
44 = 13 + 31
46 = 17 + 29
48 = 17 + 31
50 = 19 + 31
52 = 23 + 29
54 = 17 + 37
56 = 19 + 37
58 = 17 + 41
60 = 23 + 37
60 =  7 + 53
64 = 23 + 41
64 =  3 + 61
66 = 13 + 53
68 =  7 + 61
70 = 11 + 59
72 = 19 + 53
74 = 37 + 37
74 =  3 + 71
76 =  5 + 71
78 = 19 + 59
78 =  5 + 73
80 =  7 + 73
82 = 11 + 71
84 = 31 + 53
84 = 11 + 73
86 = 13 + 73
88 = 29 + 59
90 = 37 + 53
90 = 19 + 71
92 = 31 + 61
94 = 47 + 47
94 =  5 + 89
96 = 23 + 73
98 = 37 + 61
100 = 41 + 59
100 =  3 + 97
36 =  7 + 29
40 = 17 + 23
42 = 13 + 29
44 =  7 + 37
46 =  5 + 41
48 = 11 + 37
50 = 13 + 37
52 = 11 + 41
54 = 13 + 41
56 = 13 + 43
58 = 11 + 47
60 = 19 + 41
62 = 31 + 31
64 = 17 + 47
66 = 29 + 37
66 =  7 + 59
70 = 29 + 41
70 =  3 + 67
72 = 13 + 59
74 = 31 + 43
76 = 29 + 47
76 =  3 + 73
78 = 17 + 61
80 = 37 + 43
82 = 41 + 41
82 =  3 + 79
84 = 23 + 61
84 =  5 + 79
86 =  7 + 79
88 = 17 + 71
90 = 31 + 59
90 = 17 + 73
92 = 19 + 73
94 = 41 + 53
96 = 43 + 53
96 = 17 + 79
98 = 31 + 67
100 = 29 + 71
END OF PROBLEM
END PROGRAM
```

It should be noted, however, that this problem may be solved by hand (or using an adding machine).

Excursion 10—Euclid's Primes

A *prime* is a positive integer N greater than 1 that cannot be written as the product of two positive integers unless one of the factors is also N. Positive integers that can be written as the product of smaller positive integers are called *composite*. The integer 1 is neither prime nor composite, but is called a *unit*.

Thus 2, 3, 5, 7, 11, 13, 1009, and 7432339208719 are a few examples of primes, while 6, 38, 221, and 1003 (which equals $17 \cdot 59$) are composite. Euclid (*Elements*, Book IX) proved that there are infinitely many primes. There are also infinitely many composite integers. (Why?)

It is easy to tell whether a given integer is even or odd, but to determine whether a given positive integer is prime or composite is by no means simple if the integer is large. Actually, we *do* know how to determine whether or not a given positive integer is prime; it is just that for large integers there is not time enough to carry out the simple algorithm even using the fastest modern computers. The algorithm itself is simple. In its crudest form it is:

Given a positive integer $N > 1$, if there exists an integer D, with $1 < D < N$ which divides N, then N is composite, otherwise N is prime.

An easy refinement shows that there is really no need to test *all* of the integers D between 1 and N—we need only test the *primes*. Furthermore, since, if N has a factor $> \sqrt{N}$, then it must also have a factor $< \sqrt{N}$, we need test only those primes $\leq \sqrt{N}$. We, thus, have a much faster test algorithm:

If some prime $D \leq \sqrt{N}$ divides a positive integer $N > 1$, then N is composite, otherwise N is prime.

Even this test was so time-consuming that it was 1964 before it was shown that $N = 2^{11213} - 1$ is prime. This result was needed in research related to Mersenne primes and perfect numbers.

It is almost obvious that if one takes a multiple of 3 and adds 1 to it, the result is *not* divisible by 3. A remainder of 1 results. In a similar fashion if a multiple of 5 is increased by 1, the result is not divisible by 5. In general terms, an integer that is 1 more than a multiple of K is not divisible by K.

In attempting to prove that the number of primes is infinite, perhaps the most natural way to attack the problem would be to produce a formula that, given a prime, will produce the next larger prime. This, however, has not been possible. Even today, no one has developed a *formula* that will produce the "next larger prime," in spite of many attempts to do so. The pattern of primes is irregular in the extreme. Euclid realized that it was unnecessary to produce the *next* larger prime and that it would be sufficient merely to show the existence of *some* prime that was larger than the supposedly largest prime to show that the sequence of primes has no end.

Euclid's proof that the number of primes is infinite uses this observation. He considers

$(2\cdot3) + 1 = 7$ is not divisible by either 2 or 3,
$(2\cdot3\cdot5) + 1 = 31$ is not divisible by 2 or 3 or 5,
$(2\cdot3\cdot5\cdot7) + 1 = 211$ is not divisible by 2 or 3 or 5 or 7,
$(2\cdot3\cdot5\cdot7\cdot11) + 1 = 2{,}311$ is not divisible by 2 or 3 or 5 or 7 or 11,
$(2\cdot3\cdot5\cdot7\cdot11\cdot13) + 1 = 30{,}031$ is not divisible by 2 or 3 or 5 or 7 or 11 or 13,

.
.
.

$(2\cdot3\cdot5\cdot7\cdot\ldots\cdot P_K) + 1$, is not divisible by any prime $\leq P_K$.
where P_K is the Kth prime

The proof follows:

Assume there exists a largest prime P_K. Add 1 to the product of all the primes $\leq P_K$

$$N = 1 + (2\cdot3\cdot5\cdot7\cdot11\cdot13\cdot\ldots\cdot P_K).$$

Since N is greater than 1, N is not a unit and must be either prime or com-

posite. However, N has no prime factor $\leq P_K$, since each such factor produces a remainder of 1 upon division. Hence, either N is a prime greater than P_K or N is composite, in which case each of its prime factors must be greater than P_K. In either case, the existence of a prime larger than the supposed largest prime P_K has been demonstrated. Hence, the sequence of primes is unending (i.e., infinite).

Actually, there are values of P_K for which the corresponding N is really prime (for example, if $P_K = 3$, then $N = 1 + 2 \cdot 3 = 7$ is prime). There are also values of P_K such that the corresponding value of N is composite (with all its factors $> P_K$, of course). Your problem is to determine the smallest value of P_K such that $N = 1 + (2 \cdot 3 \cdot 5 \cdot \ldots \cdot P_K)$ is composite.

The reader interested in exploring more about primes will find a wealth of material in almost any text on number theory (Dewey decimal number 512, Library of Congress Number QA 241). Two conjectures that can be investigated using techniques available to the reader follow:

CONJECTURE. *Given an integer N, there exists a sequence of consecutive integers, each of which is composite (or as a special case, find a sequence of 2,000 consecutive integers containing no primes).*

CONJECTURE. *The sequence of integers of the form*

$$\{3n + 1\} = \{4, 7, 10, 13, 16, 19, 22, \ldots\}$$

contains infinitely many primes.

Both of the above conjectures are true, but the fact that someone else has already discovered a proof of them does not detract from your credit if you, too, can do so. Techniques related to those used by Euclid will produce the desired results, although neither of these conjectures was considered by Euclid (as far as we can determine).

Actually, the sequence of integers of the form $\{3n - 1\}$ also contains infinitely many primes, but the methods of proof we have seen are not related to those discussed in this section. A general theorem states that the arithmetic progression (sequence)

$$\{a + nb\} = \{a, a + b, a + 2b, a + 3b, a + 4b, \ldots\}$$

will contain infinitely many primes if the integers $a > 0$ and $b > 0$ are relatively prime (i.e., have no common factors greater than 1). This is a hard theorem to prove, in general, but many special cases ($\{3n + 1\}$, $\{4n - 1\}$, e.g.) can be proved by methods similar to those discussed above.

Excursion 11—Powers of 2

Look at a table of powers of 2. Can you detect the pattern 2, 4, 8, 6, 2, 4, 8, 6, . . . in the final (units) digits? Now can you *prove* that this pattern will continue throughout the table of 2^N no matter how far it goes? If you have studied modular (clock) arithmetic the mod 10 system will be a big help. If not, just think about what happens to the final digits as you multiply by 2.

Now let's hunt for a pattern in the last *two* digits of the powers of 2. What about the powers of 3, or 4, or 5. . . .? What about a pattern in the last three

digits? There are a lot of conjectures to be made, and then proved or disproved.

Let us also look at the beginning digits of the powers of 2. No pattern is apparent.

However, mathematicians can prove that given any sequence of digits $S = d_1d_2d_3 \ldots$, there is an integer N_S such that 2^{N_S} begins with the sequence of digits S. For example, if $S = 81$ then $2^{13} = 8192$ and $N_S = 13$. Our task is to find the *smallest positive* N_S associated with $S = 1, 2, 3, 4, \ldots, 25$. See figure 7.5.

For S =	1	2	3	4	5	6	7	8	...	25
N_S = smallest positive integer such that 2^{N_S} starts with the digits S	4	1	5	2	9	6	46	3	...	?
Value of 2^{N_S}	16	2	32	4	512	64	70368744177664	8	...	?

Fig. 7.5

Actually, the property discussed above holds not only for powers of 2 but also for powers of any positive integer that is not a power of 10 (i.e., $b \neq 0, 1, 10, 100, \ldots$). Pick some one-digit number other than 0, 1, or 2 and find the smallest power N_S such that b^{N_S} begins with S for $S = 1, 2, 3, \ldots$.

BIBLIOGRAPHY

There are many texts on theory of numbers. They are often cataloged under Dewey 512.18 or Library of Congress QA 241–48. Consult your own library. A few particularly suitable volumes are the following:

Andree, Richard V. *Selections from Modern Abstract Algebra.* 2d ed. New York: Holt, Rinehart & Winston, 1971.

Barnett, I. A. *Some Ideas about Number Theory.* Washington, D.C.: National Council of Teachers of Mathematics, 1961.

Davenport, Harold. *The Higher Arithmetic.* New York: Harper & Row, Torchbooks, 1960.

Dynkin, E. B., and V. A. Uspenskii. *Problems in the Theory of Numbers.* Lexington, Mass.: D. C. Heath & Co., 1963.

Jones, B. W. *The Theory of Numbers.* New York: Holt, Rinehart & Winston, 1955.

Maxfield, John E., and Margaret W. Maxfield. *Discovering Number Theory.* Philadelphia: W. B. Saunders Co., 1972.

Niven, I. M., and H. S. Zuckerman. *An Introduction to the Theory of Numbers.* New York: John Wiley & Sons, 1960.

Shanks, Daniel. *Solved and Unsolved Problems in Number Theory.* New York: Spartan Books, 1962.

8

Non-Euclidean Geometries

Bruce A. Mitchell

It is the purpose of this paper to introduce the reader to the rather strange world of non-Euclidean geometry. The introduction begins with the historical events that precipitated a closer look at Euclid's fifth postulate and the consequential development of the non-Euclidean geometries. In addition to the discussion of the fifth postulate, the emergence of three additional subtopics (absolute geometry, hyperbolic plane geometry, and elliptic plane geometry) should be obvious. Some of the major results will be presented with the idea that interested students will pursue the details in the appropriate references presented in this paper.

The Fifth Postulate of Euclid

When Euclid organized his geometry in the *Elements*, he began with five "common notions" and five postulates (Wolfe, p. 4). It is the fifth postulate of Euclid that has been the center of so much controversy. Although there are many equivalent forms of this postulate (Eves and Newsom, pp. 53–54; Wolfe, pp. 25–26), the particular one that is usually used in place of the rather lengthy original version is credited to John Playfair (1748–1819).

PLAYFAIR'S AXIOM. *Through a given point can be drawn only one line parallel to a given line.*

The troublesome part of this postulate is that to accept it, one must place a certain amount of faith that every other line through the given point will eventually intersect the given line. Suppose another line through the given point meets the parallel at an angle of 179.9999999999999 degrees. Can we really be sure it intersects the given line? Now, if someone could deduce the fifth postulate from the others, it could be classified a theorem. Assuming the acceptance of the other postulates, Euclidean geometry would then be the "true" geometry.

Because of a strong belief of the early mathematicians in the correctness of Euclidean geometry, there were numerous attempts to prove the fifth postulate. Proclus (410–485) tells of Ptolemy's attempt, and Proclus himself attempted a proof. Some of the attempts that followed were made by Nasiraddin (1201–1274), Wallis (1616–1703), Saccheri (1667–1733), Lambert (1728–1777), and Legendre (1752–1833). Generally speaking, the attempts all failed. Somewhere in each attempted proof, a theorem that was an equivalent form of the fifth postulate, or a tacit assumption about parallels, was used. Today we know that it is impossible to prove the fifth postulate from the others; it is independent. For those who are interested in exploring in more detail these attempts, Wolfe's book, pp. 26–41, does an excellent job. Details about the fifth postulate and historical developments are given good accounts in Bell; Bergamini; Courant and Robbins; Eves; Fawcett; *Insights into Modern Mathematics;* Moise; *Mathematics in the Modern World;* and Wolfe.

Absolute Geometry

It was the attempt by Saccheri to establish Euclid's fifth postulate that led directly to absolute geometry. If Saccheri had not believed so strongly in the correctness of the Euclidean fifth postulate, he might have had the distinction of discovering non-Euclidean geometry.

Let us follow a little of what Saccheri did. It turns out the first twenty-eight propositions (theorems) proved by Euclid in the *Elements* were proved without using the fifth postulate. (See Wolfe, pp. 220–22, for a listing of the propositions.) Perhaps Euclid himself was trying to avoid what he thought might be a controversial postulate. (This is pure conjecture.) At any rate, since none of the proofs of these first twenty-eight propositions depends on the fifth postulate, they are available for our use in any system that assumes the other postulates. This is exactly what Saccheri did. Using the other postulates and the first twenty-eight propositions, Saccheri examines what he called the "isosceles birectangle." Don't let that name scare you because it is really quite simple!

DEFINITION. *Figure ABCD is an isosceles birectangle, usually called a Saccheri quadrilateral, if* $\overline{AB} \perp \overline{BC}$, $\overline{DC} \perp \overline{BC}$, *and* $AB = DC$. (See fig. 8.1.)

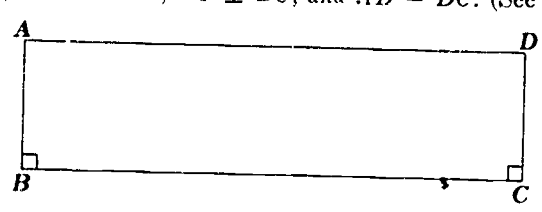

Fig. 8.1

Problem. Using Euclid's common notions, all his postulates except the fifth, and any of the first twenty-eight propositions, prove that given a Saccheri quadrilateral, as defined and pictured, $AC = BD$ and $\angle A = \angle D$. (This is very easy—try it.)

Although it can be easily demonstrated that $\angle A = \angle D$, without the use

of Euclid's fifth postulate no conclusion can be made about the measure of these two angles. Saccheri considered three possibilities:

1. Both $\angle A$ and $\angle D$ could be right angles.
2. Both $\angle A$ and $\angle D$ could be acute angles.
3. Both $\angle A$ and $\angle D$ could be obtuse angles.

At this point Saccheri attempted to rule out 2 and 3 as possibilities. If he had been successful in doing this, he would have established possibility 1 as correct. Since 1 is equivalent to the fifth postulate, Saccheri would have established Euclidean geometry as a system free from conjecture about the fifth postulate!

Although his arguments were not correct, in the process of their development Saccheri established many of the classical theorems of non-Euclidean geometry. (See Eves and Newsom, pp. 55–59; Kattsoff, pp. 630–36; Fawcett and Cummins, p. 273.)

We are led quite naturally by this discussion into absolute geometry. The geometry that results by ignoring the fifth postulate and deducing theorems without its use is absolute geometry. Saccheri, then, with his assault on the hypothesis of the acute and obtuse angles proved a substantial number of theorems that do not require the use of the fifth postulate or its equivalents. These theorems together with the first twenty-eight propositions of Euclid constitute the core of absolute geometry. Five of the more important results appear in Eves and Newsom, p. 58. On pp. 125–31 of Moise's *Elementary Geometry from an Advanced Standpoint*, there is a readable treatment of the theorems that can be proved in absolute geometry, along with their proofs. Two particularly interesting results are these (Kattsoff):

1. In any Saccheri quadrilateral the upper base (summit) is equal in length to or longer than the lower base.
2. The sum of the interior angles of a triangle is always equal to or less than 180.

Hyperbolic Geometry

The three men usually credited with the discovery of non-Euclidean geometry are Carl Gauss (1777–1855), Johann Bolyai (1802–1860) and Nicolai Lobachevski (1793–1856). Gauss and Bolyai were probably the first to see the independence of the fifth postulate, but Lobachevski was the first to publish an organized development of the subject. The geometry that these men were studying was hyperbolic non-Euclidean geometry and is sometimes referred to as Lobachevskian geometry. There are many fascinating results in this geometry. If you share this feeling after the cursory examination given here, Wolfe's *Non-Euclidean Geometry* is an excellent readable source for further study.

As in the case of absolute geometry, the first nine postulates (five common notions and four postulates) are available as well as the first twenty-eight propositions. (Why?) In hyperbolic geometry, instead of the fifth postulate's

being ignored, it is replaced by a contradictory alternative that will be called the characteristic postulate of hyperbolic plane geometry.

CHARACTERISTIC POSTULATE OF HYPERBOLIC PLANE GEOMETRY. *Through a given point not on a given line there exists more than one line that does not intersect the given line.*

Before looking at the theorems of hyperbolic geometry, it should be mentioned that indirect proof is used in many cases, and, at the beginning, reference is made to Pasch's axiom. Put very briefly, the spirit of indirect proof involves assuming the negation of the entire statement that is being proved and, as a consequence of this assumption, finding a contradiction of a known fact. Since the assumption led to a contradiction, it is an incorrect assumption; therefore, its negation, which is the original statement, has been proved within the structure of the system (Van Engen, pp. 640–42). Euclid made a few tacit assumptions in developing his geometry (Wolfe, pp. 5–9). Moritz Pasch (1843–1930) recognized one of these. The result of this was Paschs' axiom, which roughly states:

If a line passes through one side of a triangle, not at a vertex, then it must intersect at least one of the other two sides.

It follows from this that if a line enters a triangle at a vertex, then it intersects the opposite side (Kattsoff, pp. 630–31).

Because of the characteristic postulate of hyperbolic geometry, three classes of lines are generated (given line l and a point P not on l) as described and pictured in figures 8.2 through 8.4.

Case 1, shown in figure 8.2: those lines through P that intersect l.

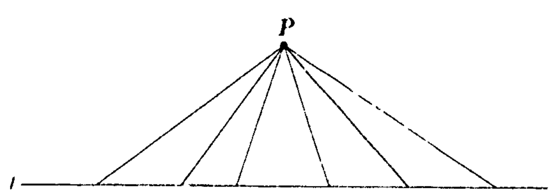

Fig. 8.2. Case 1

Case 2, shown in figure 8.3: those lines that are the "first" lines on the left and right of intersecting line $\overline{PQ}$ that do not intersect l.

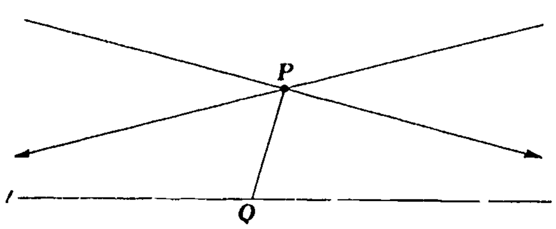

Fig. 8.3. Case 2

Case 3, shown in figure 8.4: those lines through P contained within the angles formed by the lines in case 2.

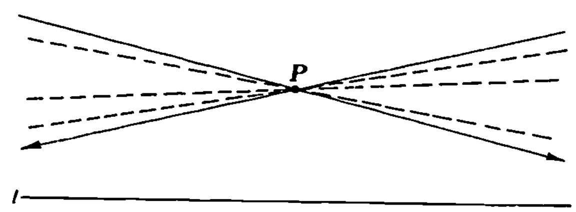

Fig. 8.4. Case 3

The lines in case 1 are called *intersecting,* the lines in case 2 are called *parallel* (it is important to note that parallel lines in hyperbolic geometry refer to the first nonintersecting lines), and the lines in case 3 are called *nonintersecting* lines. For a discussion of case 2 see Wolfe, pp. 66–67; or Moise, pp. 306–8. It must be proved that the lines described in case 3 do not intersect the given line; that proof is included in the outline of the proof of the first theorem. We are ready to summarize some of the above discussion in a theorem.

THEOREM. *If l is any line and P is a point not on l, then there are always two lines through P which* do not intersect *l and (1) which make equal acute angles with the perpendicular from P to l; and (2) every line through P lying within the angle containing that perpendicular and one of the two nonintersecting lines described in (1) intersects l, while (3) every other line through P does not.*

Establish the fact (using the characteristic postulate and Dedekind's theorem) that the lines through P are divided into two sets: (1) those that intersect l and (2) those that do not intersect l. Show next that the dividing lines of the two sets are the first of those lines, one on the "left" and one on the "right," that do not intersect l. Parts 1, 2, and 3 of the theorem refer to these two lines. (See fig. 8.5.)

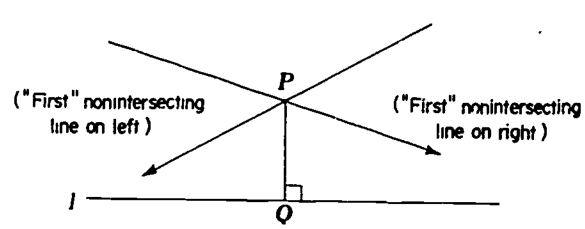

Fig. 8.5

DEDEKIND'S THEOREM. *Let A and B be sets of real numbers such that—*

1. *every real number is either in A or in B;*
2. *no real number is in A and in B;*
3. *neither A nor B is empty;*
4. *if $a \in A$ and $b \in B$, then $a < b$.*

Then there is one and only one real number c such that $a \leq c$ for all $a \in A$ and $c \leq b$ for all $b \in B$.

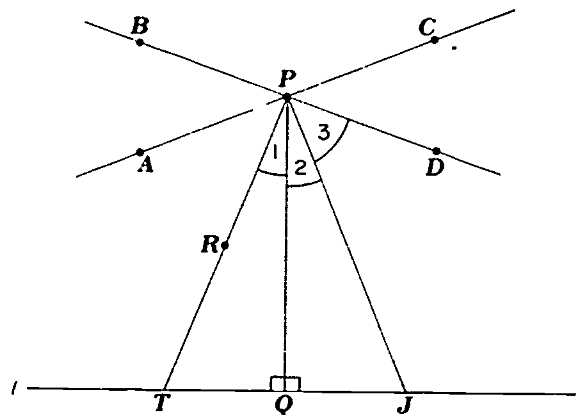

Fig. 8.6

1. Assume $\angle APQ \neq \angle DPQ$, say $\angle APQ$ is larger. Construct $\angle RPQ = \angle DPQ$. $\overline{PR}$ intersects l, say at T. (Why?) Construct $JQ = TQ$. (See fig. 8.6.) $\triangle$ [illegible]$PQ \cong \triangle JPQ$. (Why?) $m\angle 1 = m\angle 2$, but $m\angle 1 = m\angle 2 + m\angle 3$ with [illegible]rger than zero. This means $m\angle 2 = m\angle 2 + m\angle 3$ with $m\angle 3 > 0$, and th[illegible] [illegible]npossible. (The outlines of the proofs [illegible] this paper are lacking in detail and rigor. It is the intent that the reader can supply these.) The demonstration that the angles are acute is left to the reader.

2. Use the fact that the lines in (1) are the first nonintersecting lines.

3. Assume $\overleftrightarrow{PR}$ intersects l at J. PQL is a triangle with $\overleftrightarrow{PD}$ passing through the vertex. $\overleftrightarrow{PD}$ must intersect l by the theorem proved from Pasch's axiom. Since $\overleftrightarrow{PD}$ was given as nonintersecting, we have a contradiction, and the assumption is incorrect. (See fig. 8.7.)

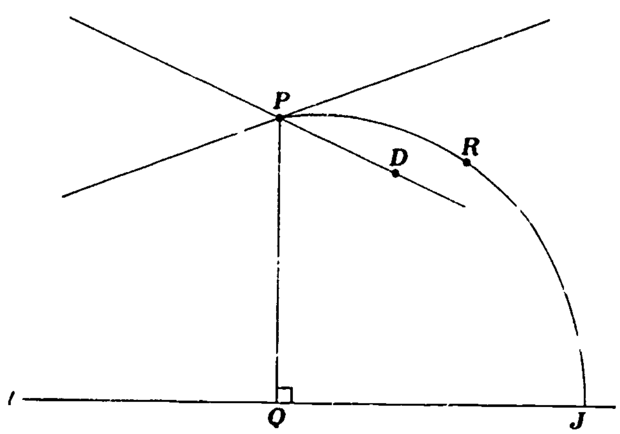

Fig 8.7

Problem: Which of the three classes of lines described above would the Euclidean "parallels" fit?

In order to explore, briefly, the idea of the ideal triangle, a definition of ideal point is necessary.

DEFINITION. *If two lines are parallel (first nonintersecting lines), then they share an ideal point,* Ω *(omega)*

DEFINITION. *Given two ordinary points A and B and the ideal point* Ω, *the figure* $AB\Omega$ *is an ideal triangle with* $A\Omega$ *parallel to* $B\Omega$.

After showing that the exterior angle of an ideal triangle is greater than the opposite interior angle at the ordinary point, the topic of congruent ideal triangles is developed (Wolfe, pp. 73–76).

It is now time to get back and see what's happening to the Saccheri quadrilateral in this geometry. We shall do this via a theorem. (See fig. 8.8.)

THEOREM. *If ABCD is a Saccheri quadrilateral with right angles at band C, then (1) the angles at A and D are equal. (2) the angles at A and D are acute, and (3) the line connecting the midpoints of the base and summit is perpendicular to each.*

(1) and (3) have already been proved in absolute geometry, and the same proofs can be used here.

(2) There exists a parallel to $\overline{BC}$ through A and D, $A\Omega$ and $D\Omega$. Since $AB = DC$ and $\angle ABC = \angle DCE$, $\triangle AB\Omega \cong \triangle DC\Omega$ by a congruence theorem for ideal triangles. Now, $\angle 2 = \angle 5$ by corresponding parts; $\angle 1 > \angle 4$, since $\angle 1$ is an exterior angle of $\triangle AD\Omega$. Therefore $\angle 1 + \angle 2 > \angle 4 + \angle 5$. $\angle 4 + \angle 5 = \angle 3$, so $\angle 1 + \angle 2 > \angle 3$. This makes $\angle 3$ acute!

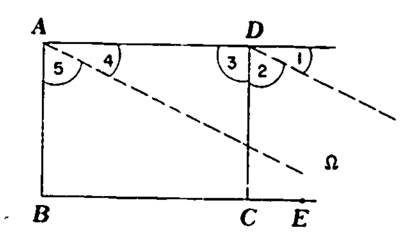

Fig. 8.8

Another strange quadrilateral is the Lambert quadrilateral.

DEFINITION. *A quadrilateral with three of its interior angles right angles is a Lambert quadrilateral.*

What seems a little unusual about this quadrilateral is the fact that it can be proved in hyperbolic geometry that the fourth angle is acute! (Wolfe, pp. 79–80.)

THEOREM. *The sum of the interior angles of every right triangle is less than 180.*

Construct $\angle 3 = \angle B$. At M, the midpoint of hypotenuse $\overline{AB}$, construct $\overline{MP} \perp \overline{CB}$. Mark off $AQ = PB$. (See fig. 8.9.) Now $\triangle QAM \cong \triangle PBM$. (Why?) $\angle 1 = \angle 2$, and P,M,Q are collinear (why?), and $ACPQ$ is a Lambert

quadrilateral (why?), with $\angle CAQ$ acute Since $\angle 3 + \angle 4 < 90$ and $\angle B = \angle 3$, this means $\angle 4 + \angle B < 90$ With $\angle C = 90$, $\angle C + \angle 4 + \angle B < 180$.

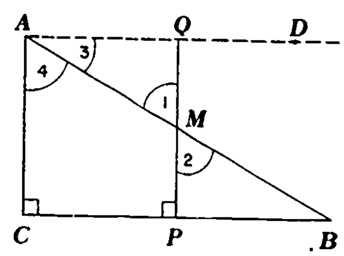

Fig. 8 9

Topics that follow in a more thorough study are properties of ordinary triangles, nonintersecting lines, construction of parallels, angle defect, area, and models—not necessarily in that order. Problems can be found in Wolfe's book, in Maiers's article, in most textbooks on non-Euclidean geometry, and in Eves and Newsom. One that I will leave you with is important in the proof of (2) for a Saccheri quadrilateral. Are the base and summit of a Saccheri quadrilateral intersecting, parallel, or nonintersecting? Which is greater, base or summit? It might be finally interesting to note that in hyperbolic geometry two triangles with all of the angles in one equal respectively to those in the other are congruent!!

Elliptic Geometry

In 1854 Bernhard Riemann delivered a lecture dealing with the boundedness and infinitude of a line that made a major contribution to geometry. Any two straight lines in a plane intersect if the first, second, and fifth postulates of Euclid are modified to the following:

1'. Two distinct points determine at least one straight line.
2'. A straight line is unbounded
5'. Two straight lines always intersect each other.

Another geometry, equally consistent to Euclidean and hyperbolic, results.

An interesting result that comes early in this geometry is that all perpendiculars erected on the same side of a given line l are concurrent in a point P called the *pole* (Wolfe pp. 175–76).

It is also true that $PA = PB = PC = PD = q$. (See fig. 8.10.)

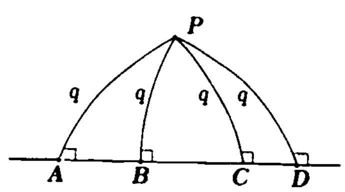

Fig. 8.10

Problem. Given figure 8.11 with $\overline{OA} \perp \overline{AB}$, $\overline{OB} \perp \overline{AB}$, extend $\overline{OA}$ to O' such that $AO' = AO$. Draw $\overline{O'B}$ and show that O', B, and O are collinear.

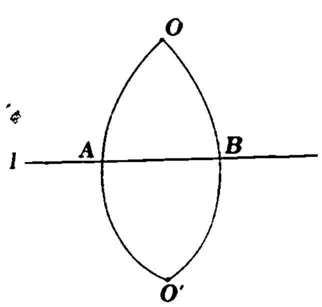

Fig. 8.11

The figure formed is called a *digon*. In this geometry it is possible for two of the same kind of line Euclid was talking about to enclose space.

THEOREM. *Given a triangle ABC with a right angle, if one of its sides is less than q (the constant perpendicular distance from the pole to the line), the angle opposite that side is acute.* See figure 8.12.

Since O is the pole, the angle at A is a right angle ($\angle CAO$), with $\angle CAB < \angle CAO$ and therefore acute.

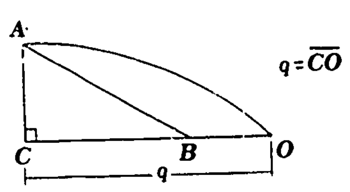

Fig. 8.12

THEOREM. *The line joining the midpoints of the base and summit of a Saccheri quadrilateral is perpendicular to both of them, and the summit angles are equal and obtuse.* See figure 8.13.

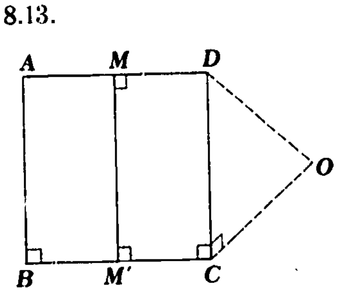

Fig. 8.13

Again the proof of the midpoint line and equal summit angles is unchanged. To prove that the summit angles are obtuse, extend AD and BC and they

will meet at a pole, say O. (Why?) $\angle DCO$ is a right angle with CO shorter than q. By the theorem above, $\angle CDO$ is acute. This means that $\angle CDM$ is obtuse, and the theorem is established.

It follows that in a Lambert quadrilateral the fourth angle is obtuse, and the sum of the interior angles of a right triangle is greater than 180. The proof for the angle sum theorem is very similar to the corresponding proof in hyperbolic and can be supplied by the reader. Can you generalize the interior angle sum theorems to a general triangle? A quadrilateral?

There are models on which the properties described for hyperbolic and elliptic have physical justifications (Moise, pp. 114–21; Eves and Newsom, pp. 68–72; Groza, pp. 287–89; Courant and Robbins, pp. 221–25; *Mathematics in the Modern World*, pp. 118–19); but that is a whole topic in itself.

The author has made an attempt to get the reader involved just deeply enough in the geometries to let his curiosity get the b[illegible] f him in at least one of the topics discussed. An adequate list of references [illegible] een supplied to help satisfy the curiosity. I hope that many of you will do this sort of follow-up work on one of the topics. I think you will find it fascinating!

REFERENCES

Bell, E. T. *Men of Mathematics*. New York: Simon & Schuster, 1937.

Bergamini, David, and the Editors of *Life*. *Mathematics*. New York: Time, 1963.

Courant, Richard, and Herbert Robbins. *What is Mathematics?* New York: Oxford University Press, 1941.

Coxeter, H. S. M. *Introduction to Geometry*. New York: John Wiley & Sons, 1961.

Eves, Howard. *A Survey of Geometry*, vol. 1. Boston: Allyn & Bacon, 1963.

Eves, Howard, and Carroll V. Newsom. *An Introduction to the Foundations and Fundamental Concepts of Mathematics*. New York: Rinehart & Co., 1958.

Fawcett, Harold P., and Kenneth B. Cummins. *The Teaching of Mathematics from Counting to Calculus*. Columbus, Ohio: Charles E. Merrill Publishing Co., 1970

Groza, Vivian Shaw. *A Survey of Mathematics*. New York: Holt, Rinehart & Winston, 1968.

Insights into Modern [illegible]ematics. Twenty-third Yearbook of the National Council of Teachers of M[illegible]ematics. Washington, D. C.: The Council, 1957.

Kattsoff, Louis O. "The Saccheri Quadrilateral." *Mathematics Teacher* 55 (December 1962): 630-36.

Lieber, L. R. *Non-Euclidean Geometry*. Lancaster, Pa.: Science Printing Co., 1940.

Maiers, Wesley W. "Introduction to Non-Euclidean Geometry." *Mathematics Teacher* 57 (November 1964): 457-61.

Mathematics in the Modern World: Readings from "Scientific American." San Francisco: W. H. Freeman & Co., 1968.

Moise, Edwin. *Elementary Geometry from an Advanced Standpoint*. Reading, Mass.: Addison-Wesley Publishing Co., 1963.

Van Engen, Henry. "Strategies of Proof in Secondary Mathematics." *Mathematics Teacher* 63 (December 1970): 637-45.

Wolfe, Harold E. *Non-Euclidean Geometry*. New York: Holt, Rinehart & Winston, 1945.

9

Boolean Algebras

Wade Ellis

The purpose of this all too brief expository tract is to provide enough general insight into the concept and application of Boolean algebras to create and stimulate interest on the part of students who have studied some ordinary algebra—interest which, if sustained, will bring the deep and rewarding satisfaction that always accompanies productive mental labor. But the rewards should extend beyond the gratification of a natural desire, even need, for accomplishment. Mastery of the historic, recent, and continuing development of Boolean algebras will provide the student with knowledge that continues to grow, to become more beautiful, and to reveal its deeper and broader associations with an ever-increasing variety of mathematical concepts and applications.

A list of references appears at the end of the article. The books in this list provide insights into the earliest developments in the subject area as recorded by Boole himself. They also provide more extensive expository treatment than can be given here and include enough examples of applications and relations with other parts of mathematics, together with their own bibliographies, to whet our curiosities and interests.

Ordinary algebra uses the real number field. The field axioms should be reviewed carefully in connection with this article; by all means they should be at hand for careful comparison when the different sets of axioms for a Boolean algebra are encountered here. The differences and similarities between the two types of systems are defined by the sets of axioms. They are more clearly and fully explained by the bodies of theorems developed from the axioms. Thus no axiomatic system is fully understood until all the consequences of its axioms are known.

Just as the field axioms are satisfied by various fields (e.g., the complex field; the real field; the algebraic fields; the rational field; the fields of residue

classes of integers modulo p where p is a prime, or modulo p^n where p is prime, and n is a natural number, $n \geqslant 2$), so there are various systems of the type $\langle S; \oplus, \otimes \rangle$, each consisting of a set of elements and two operations jointly satisfying the Boolean algebra axioms. It should be remembered that every Boolean algebra in the sense of the preceding paragraph has all the properties that emerge as consequences of the axioms.

We begin now with ordinary familiar truth tables.

Propositional Calculus

A proposition is a sentence that is either true or false, but not both (and also not neither). Propositions abound, not only in nature, but in everyday life. We might wish that all sentences could be true, but that is not possible. The proposition $3 + 5 = 8$ is true. However, if we change any one (and only one) of the numbers 3, 5, or 8, the new proposition will be false: $3 + 6 = 8$ is false. Moreover, if we leave one of the numbers unspecified, the expression is not a proposition; $n + 5 = 8$ is a sentence, but it is neither true nor false, and therefore it is not a proposition. It *becomes* a proposition whenever n is a specified number. Not only that, but as soon as n is specified, we *know* whether the proposition is true or false. As a matter of fact, we know that for every value of n except $n = 3$ the proposition is false and that for $n = 3$ the proposition is true.

A sentence that is not a proposition but becomes a proposition when exactly one unspecified word (repeated or not) in it is specified is called a propositional function. A convenient example is

$$\frac{n}{2} = \left[\frac{n}{2}\right]$$

which says, remembering that $\left[\frac{n}{2}\right]$ is the symbol for "the largest integer in $\frac{n}{2}$," precisely that "n is an integer that is divisible by 2." If we specify that $n = 6$, the sentence becomes a proposition that is true. Similarly for $n = 8$. Clearly the possibilities are boundless. However, if we specify that $n = 9$, the sentence becomes a proposition that is false. Similarly for $n = 11$. Again the possibilities are boundless.

Let us now resort to the use of the familiar notation for a function. Let p be the name of our function, and let its domain be the set I of all integers, positive, negative, and zero. Let n be the generic name of the elements of the domain. Let the range of p be the set $R = \{T, F\}$, $T \equiv$ (the proposition is true), $F \equiv$ (the proposition is false)}. Then p maps I onto R through

$$p: I \to R, \; p(n) = \begin{cases} T \text{ if } \frac{n}{2} = \left[\frac{n}{2}\right] \\ F \text{ if } \frac{n}{2} \neq \left[\frac{n}{2}\right] \end{cases}$$

A short table of values for this function follows.

n	$p(n)$
6	T
8	T
9	F
11	F

This table can be extended without bound by simply entering values of n and computing the corresponding values of p according to its defining formula.

Another convenient example is provided if we map I onto R through a similar definition of a function q:

$$q: I \rightarrow R,\ q(m) = \begin{cases} T \text{ if } \frac{m}{3} = \left[\frac{m}{3}\right] \\ F \text{ if } \frac{m}{3} \neq \left[\frac{m}{3}\right] \end{cases},$$

where now m is the generic name of the elements of I. A short table of values for the function q shown below:

m	$q(m)$
6	T
8	F
9	T
11	F

Again the table can be extended indefinitely, as in the case of p. Let us keep these functions and tables in mind as we proceed.

The functions have immediate but rather simple-minded applications. Suppose we have one device that can determine whether an integer is congruent to zero modulo 2, that is, for a given n, whether $p(n) = T$ or $p(n) = F$. Let this device be connected to a switch controlling lighted signals so that if the switch is closed ($p(n) = T$), the T signal is lighted and if the switch is open ($p(n) = F$), the F signal is lighted. A simple toy, suitable for use in a kindergarten, could now be constructed. But before we consider that, let's drag in some other ideas.

Suppose we feed through the device in succession 1,000 *random numbers* and count the T's and F's as they light up. How many T's would be expected by the time the process was complete? Suppose the number actually found should be much different from what was expected. What conclusion(s) could be reached? (Don't jump to conclusions! Nor away from them! Your hypothesized, observed discrepancy could be due to one reason or any combination of several reasons.) It would be entertaining—a prime reason for studying and doing mathematics—but perhaps distracting to continue asking and answering such questions. These few are included only for the purpose of providing a glimpse into the kind of experiences the mind provides even in the absence of hardware. And if you want hardware, it is much less expensive to explore the realm mentally before even beginning to design machines or draw construction diagrams, to say nothing of starting to build.

It is left as a simple exercise for the reader to examine the function q as we have just examined p. Now that that's finished, let us consider working with p and q jointly. To do this in the way that will lead us toward Boolean algebra, we first make $m = n$ and use n as our symbol. We have two obvious choices

suggested by simplest joint uses of two switches. Simple diagrams will perhaps help.

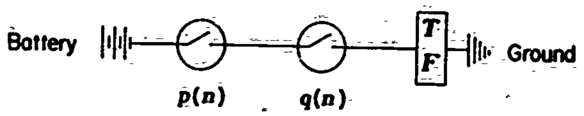

Fig. 9.1. Series connection

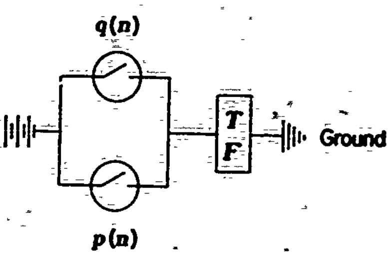

Fig. 9.2. Parallel connection

It will be interesting to use our previous tables to represent these diagrams:

n	$p(n)$	$q(n)$	Series connection	Parallel connection
6	T	T	T	T
8	T	F	F	T
9	F	T	F	T
11	F	F	F	F

The reader should verify all these entries by speculating about the various possible combinations of conditions of the switches in the diagrams. We note immediately that any integer n will generate unambiguously exactly one of the horizontal rows of entries within the entire table. Also, no integer will generate any row of entries except one of these four. We can dispense with both the specific definitions of the functions p and q and also with the integers. This is a peculiar abstraction in which we free the functions and their common domain but preserve intact their common range. We have put ourselves in position not only to complete the usual truth table entries for $p; q; \sim p; p \wedge q; p \vee q; p \Rightarrow q; p \Leftrightarrow q; p \vee (\sim p)$ (tautology); and $p \wedge (\sim p)$ (contradiction) but also the other seven columns of entries almost never listed and even more rarely used: $\sim (p \wedge q); \sim (p \vee q); p \Leftarrow q; \sim (p \Rightarrow q); \sim q; \sim (p \Leftarrow q)$; and $\sim(p \Leftrightarrow q)$. It is easily seen that each of these sixteen columns defines a function that is a "composite" of p and q; whose domain (which need not be otherwise specified) may be taken as the common domain of p and q. The reader will easily construct the entire sixteen-column table and will use the sixteen names $(p, q, \sim p, \ldots, \sim(p \Leftrightarrow q))$ of functions above to properly label the columns. The more energetic and resourceful reader will construct wiring diagrams, similar to those above, to simulate or represent these functions based on p and q. The persistent and affluent reader will even construct or assemble bits of hardware for concrete illustration.

Since the table of entries is now *complete*, with four rows and sixteen columns, it is virtually impervious. The labels on the columns can be rearranged in precisely 4! (no. not 16!) different ways without changing its import. More important, however, is the fact that the table can be represented by 16 different matrices. They are:

$C \equiv p \wedge (\sim p)$

$p \backslash q$	F	T
F	F	F
T	F	F

$p \wedge q$

$p \backslash q$	F	T
F	F	F
T	F	T

$\sim(p \Rightarrow q)$

$p \backslash q$	F	T
F	F	F
T	T	F

$\sim(p \Leftarrow q)$

$p \backslash q$	F	T
F	F	T
T	F	F

$\sim(p \vee q)$

$p \backslash q$	F	T
F	T	F
T	F	F

p

$p \backslash q$	F	T
F	F	F
T	T	T

q

$p \backslash q$	F	T
F	F	T
T	F	T

$p \Leftrightarrow q$

$p \backslash q$	F	T
F	T	F
T	F	T

$\sim(p \Leftrightarrow q)$

$p \backslash q$	F	T
F	F	T
T	T	F

$\sim q$

$p \backslash q$	F	T
F	T	F
T	T	F

$\sim p$

$p \backslash q$	F	T
F	T	T
T	F	F

$p \vee q$

$p \backslash q$	F	T
F	F	T
T	T	T

$p \Leftarrow q$

$p \backslash q$	F	T
F	T	F
T	T	T

$p \Rightarrow q$

$p \backslash q$	F	T
F	T	T
T	F	T

$\sim(p \wedge q)$

$p \backslash q$	F	T
F	T	T
T	T	F

$T = p \vee (\sim p)$

$p \backslash q$	F	T
F	T	T
T	T	T

If we replace each F by O, and each T by I, these arrays are, in order:

$C \equiv p \wedge (\sim p)$

$p \backslash q$	O	I
O	O	O
I	O	O

$p \wedge q$

$p \backslash q$	O	I
O	O	O
I	O	I

$\sim(p \Rightarrow q)$

$p \backslash q$	O	I
O	O	O
I	I	O

$\sim(p \Leftarrow q)$

$p \backslash q$	O	I
O	O	I
I	O	O

$\sim(p \vee q)$

$p \backslash q$	O	I
O	I	O
I	O	O

p

$p \backslash q$	O	I
O	O	O
I	I	I

q

$p \backslash q$	O	I
O	O	I
I	O	I

$p \Leftrightarrow q$

$p \backslash q$	O	I
O	I	O
I	O	I

$\sim(p \Leftrightarrow q)$

$p \backslash q$	O	I
O	O	I
I	I	O

$\sim q$

$p \backslash q$	O	I
O	I	O
I	I	O

$\sim p$

$p \backslash q$	O	I
O	I	I
I	O	O

$p \vee q$

$p \backslash q$	O	I
O	O	I
I	I	I

$p \Leftarrow q$

$p \backslash q$	O	I
O	I	O
I	I	I

$p \Rightarrow q$

$p \backslash q$	O	I
O	I	I
I	O	I

$\sim(p \wedge q)$

$p \backslash q$	O	I
O	I	I
I	I	O

$T = p \vee (\sim p)$

$p \backslash q$	O	I
O	I	I
I	I	I

Just as the truth tables contained all possible columns of four symbols, each symbol being a T or an F, so this table contains all possible 2-by-2 arrays of symbols, each symbol being an O or an I.

The array for $p \wedge q$ now reminds us of the multiplication table for the field of residues of the integers modulo 2, and the one for $\sim (p \Leftrightarrow q)$ looks like the addition table for that same field. So a more thorough than usual examination and manipulation of truth tables has led us to a situation having some features reminiscent of a familiar and apparently rather simple mathematical system. Let us not be diverted!

Boolean Algebras

Consider an arbitrary set having a single element: $I = \{a\}$. The power set of I (the set of all subsets of I) is $\{\phi, I\}$, where ϕ is the empty set. The operations of set union and set intersection defined on this power set have the tables shown below.

$\cup$	ϕ	I
ϕ	ϕ	I
I	I	I

$\cap$	ϕ	I
ϕ	ϕ	ϕ
I	ϕ	I

As we know, the counterparts of these arrays *must* be found in the exhaustive set of arrays in the preceding section; they are the arrays for $p \vee q$ and $p \wedge q$, respectively. George Boole's recognition of this fact, expressed in considerably different terms, led him to lay the groundwork and make the initial constructions of what is apparently properly known now as Boolean algebra.

We can reconstruct the early stages of this development, arranged in a rationally appealing if not necessarily chronological order, in the following ashion:

First the above fact was noted in the case of a one-element set I. Then a set of axioms, not all independent, was developed to describe the set algebra for the power set of I. Then a start was made to explore the axioms, leading to the development of a body of theorems and associated statements as well as modification of the axioms themselves. Continuing throughout this development has been a consideration of and application to the case of an arbitrary set I with an arbitrary number of elements. This generalization from the earliest concept proceeds in a different manner from the development in the case of fields. There the start was with the rational field. A very interesting diversion lies here—let us not follow it!

An interesting set of mutually independent axioms for a Boolean algebra is the set of so-called Huntington postulates. It is listed here because of its aesthetic appeal and the simplicity of its use as a criterion to determine whether a given mathematical system is a Boolean algebra.

Axiomatic Definition: A mathematical system $< B; \oplus, \otimes >$ is a Boolean algebra if the following axioms are satisfied.

A_0. *Closure.* If a and b are *any* elements of S, then $a \oplus b \in B$ and $a \otimes b \in B$.

A_1. *Commutative laws.* If a and b are *any* elements of B, then $a \oplus b = b \oplus a$ and $a \otimes b = b \otimes a$. (As usual we assume that "equality" is an equivalence relation.)

A_2. *Distributive laws*. If a, b, and c are *any* elements of B, then $a \otimes (b \oplus c) = (a \otimes b) \oplus (a \otimes c)$ and $a \oplus (b \otimes c) = (a \oplus b) \otimes (a \oplus c)$. (Note that the left identity is analogous to the distributive law of ordinary algebra. The right identity is different; its analogue does not hold in ordinary algebra. Remember that the word *identity* as used here means an equality that is valid even if the letter symbols are changed in a consistent manner.)

A_3. *Zero and unit elements*. The set B has two elements ϕ and I such that for any element a of B, we have $a \oplus \phi = a$ and $a \otimes I = a$.

A_4. *Complementation*. For each element a of B, there is an element, which we shall call a', such that $a \oplus a' = I$ and $a \otimes a' = \phi$.

There are equivalent sets different from this. Examples of such sets of axioms appear in Kaye's *Boolean Systems*, chapter 4 and article 1.9.

The simplicity of these axioms is immediately apparent. Their remarkable symmetry lies only slightly under the surface but is exceedingly important and valuable. This symmetry lies in the fact that the axioms are collectively unchanged if we make two changes simultaneously:

1. Change each $\oplus$ to $\otimes$ and each $\otimes$ to $\oplus$.
2. Change each ϕ to I and each I to ϕ.

Now the beauty and the importance are plain to see. Since this symmetry lies in the axioms, it must be transmitted throughout the entire body of consequences of the axioms. If we state and prove *any* theorem in this system, the *dual* of that theorem can be stated and proved by simply making the above-listed changes in the original theorem and proof. Although some theorems are self-dual, just as are *all* these axioms, many theorems either are not self-dual or can be split so that only one half needs to be proved; the other half is then established by the principle of duality.

To demonstrate how this works, let us prove the following theorem (laws of tautology).

THEOREM 1. *If B is any Boolean algebra, and if a is any element of B, then $a \oplus a = a$ and $a \otimes a = a$.*

As it stands, this theorem is self-dual. We can split it in half by omitting either of the identities. Let us omit the $a \otimes a = a$, and prove the other.

$$\begin{array}{rll}
(a \oplus a) \in B & & A_0 \\
(a \oplus a) = & (a \oplus a) \otimes I & A_3 \\
= & (a \oplus a) \otimes (a \oplus a') & A_4 \\
= & a \oplus (a \otimes a') & A_2 \\
= & a \oplus \phi & A_4 \\
= & a. & A_3
\end{array}$$

This completes the proof of this half. The second half, being the dual of the first, can be stated and its proof constructed by simply making the specified interchanges between $\oplus$ and $\otimes$ and between ϕ and I. Since this is true, we need not bother to write a proof. However, it is indispensable to understand the principle of duality as it applies here and to make certain that the statements which we consider as (mutually) dual are indeed (mutually) dual.

Five additional theorems, split into duals where appropriate, form a broader basis on which to develop the extensive theory that proceeds from the system

of axioms aud at the same time explains the nature of Boolean algebras. In each of these, $\langle B; \oplus, \otimes \rangle$ is any Boolean algebra and $a, b, \ldots$ are arbitrary elements of B.

THEOREM 2. $a \oplus I = I$; $\quad a \otimes \phi = \phi$.

Proof: (to be supplied by the reader.)

Proof:
$$\begin{aligned} a \otimes \phi &= \phi \oplus (a \otimes \phi) \\ &= (a \otimes a') \oplus (a \otimes \phi) \\ &= a \otimes (a' \oplus \phi) \\ &= a \otimes a' \\ &= \phi. \end{aligned}$$

THEOREM 3. $a \oplus (a \otimes b) = a$; $\quad a \otimes (a \oplus b) = a$.

Proof: (To be supplied by the reader.)

Proof:
$$\begin{aligned} a \otimes (a \oplus b) &= (a \oplus \phi) \otimes (a \oplus b) \\ &= a \oplus (\phi \otimes b) \\ &= a \oplus \phi \\ &= a. \end{aligned}$$

THEOREM 4. a' *(the complement of* a*) is unique and* $(a')' = a$.

Proof: Suppose B contains an a that has two complements a' and a'_1. Then $a \oplus a' = I$, $a \otimes a' = \phi$; and $a \oplus a'_1 = I$, $a \otimes a'_1 = \phi$.
Hence

$$\begin{aligned} a' &= a' \otimes I = a' \otimes (a \oplus a'_1) \\ &= a' \otimes a \oplus a' \otimes a'_1 = \phi \oplus a' \otimes a'_1 \\ &= a' \otimes a'_1 = a' \otimes a'_1 \oplus \phi \\ &= \phi \oplus (a'_1 \otimes a') = (a'_1 \otimes a) \oplus (a'_1 \otimes a') \\ &= a'_1 \otimes (a \oplus a') = a'_1 \otimes I \\ &= a'_1. \end{aligned}$$

Therefore a' is unique. To find $(a')'$, regard the symbol a' just as if it were a letter without decoration such as b. Then $(a')'$ is the complement of a', the parentheses being used to avoid confusion with a'', to which no meaning has been assigned. Then $(a')' \oplus a' = I$ and $(a')' \otimes a' = \phi$, as well as $a \oplus a' = I$ and $a \otimes a' = \phi$. And since the complement of a is unique, $(a')' = a$. This completes the proof. Note that in $a \oplus I = I$, $a \otimes I = a$, a is any element of B. If $a = \phi$, we have $\phi \oplus I = I$ and $\phi \otimes I = \phi$. This means that $\phi' = I$. Hence $(\phi')' = I' = \phi$. That is, the distinguished elements ϕ and I of every Boolean algebra are each the complement of the other.

THEOREM 5. $(a \otimes b)' = a' \oplus b'$; $\quad (a \oplus b)' = a' \otimes b'$.

THEOREM 6. $a \oplus (b \oplus c) = (a \oplus b) \oplus c$; $\quad a \otimes (b \otimes c) = (a \otimes b) \otimes c$.

Proofs for theorems 5 and 6 can be found in Kaye as well as in some of the other references; or perhaps the reader can construct his own.

Systems: $\langle S; \oplus, \odot \rangle$

It may be useful to note briefly several mathematical systems consisting of a set of elements and two binary operations governed by a code of laws (list of axioms). We might first notice the existence of a class of systems, even lower in a kind of hierarchy, in each of which there is a set and *one* binary operation with governing axioms. An example is the now familiar class of groups. One group is the set of integers with ordinary addition, the word,

ordinary indicating that the well-known laws of addition hold. Another is the set of integers, including zero, divisible by an arbitrary integer n, again with ordinary addition as the operation. These groups all have a countable infinity of elements. They are not to be confused with the groups of residue classes of integers modulo an arbitrary integer n with modified addition. Such groups have the same (finite) number of elements as the (absolute value of the) modulus.

If we again consider systems with two binary operations, we find a rich variety. A familiar system is the set of integers with ordinary addition (as just mentioned) and a kind of limited multiplication—limited because the set contains very few multiplicative inverses. Then there are rings and fields and a rich diversity of systems within each subclass and associated with subclasses. Discussions of some of these systems, comprehensible to the persistent reader, even the tyro, appear in such books as Birkhoff and MacLane, and Andrée. Diligent study, with an eye out for properties of one class of systems that seem similar to properties of other classes, as well as unflagging effort to understand and become familiar with individual systems, can be almost unbelievably rewarding and satisfying. Moreover, there are useful (as well as aesthetic) interpretations and applications of most systems.

In our own situation, with our truth tables set up first in the usual fashion, then in a conveniently modified arrangement, we found several interesting pairs of arrays. If the arrays representing $p \wedge q$ and $\sim(p \Leftrightarrow q)$ are used as the "multiplication" and "addition" tables, respectively, of an $\langle S; \oplus, \otimes\rangle$, where $S = \{0, 1\}$, we have a *field* of *characteristic* (and also *order*) 2. An example of such a field is the field of residue classes of integers modulo 2.

On another hand, if we take the arrays representing $p \wedge q$ and $p \vee q$ as "multiplication" and "addition" tables, respectively, of an $\langle B; \oplus, \otimes\rangle$, where $B = \{\phi, I\}$, we have a Boolean algebra. Now a Boolean algebra is very different from a field. It has different applications and different interpretations.

It should not be surprising that we can find tables defining different systems among these arrays. The fact is that all possible 2-by-2 arrays involving only two district elements are there. Hence if it is possible to define different two-element systems by using different pairs of arrays, it has to be possible to find the tables in this set. It will certainly be interesting for you to explore this matter further.

Simple Applications of a Boolean Algebra

We can obtain some simple applications for the Boolean algebra $\langle B; \oplus, \otimes\rangle$, where $B = \{\phi, I\}$, by first considering a set of light bulbs controlled by a collection of switches each with only on-off positions. The immediate source of power can be an ordinary base plug. Or our wiring configuration may be simply a part of the house (or other) wiring. As with so-called word problems, we have to assign meanings to our terms. Let S_1, S_2, . . . indicate specific switches and L_1, L_2, . . . indicate specific bulbs. In each switch, ϕ

indicates the "open" or "off" status with no current flowing; I indicates the "closed" or "on" status with (all) current flowing. In the case of each bulb, ϕ indicates that it is not lighted and I indicates that it is lighted. For a single bulb controlled by a single switch, the representation is so simple as to be perhaps confusing:

S_1	L_1	
ϕ	ϕ	Switch off, bulb not lighted
I	I	Switch on, bulb lighted

The tacit assumptions are that (1) the house current is on, (2) the switch works, and (3) the bulb is "good." Also, the case of a three-way bulb controlled by an in-socket so-called switch must be differentiated. Such a switch is a collection of two of our switches or a wiring configuration with equivalent effect.

We are free to reverse this control procedure by specifying that when the switch is on, the bulb is not lighted and when the switch is off the bulb is lighted. Why, do you suppose, don't we use this arrangement generally?

If we want a bulb to be controlled by two switches (at different locations) in such a way that the bulb's status will be reversed (from off to on *or* from on to off) by reversing the status of either switch, our representation has to provide all four possible combinations of statuses for the switches as well as both statuses for the bulb. This may be done in either the columnar or the (bordered) square array form with the L_1 entries inside:

S_1	S_2	L_1
ϕ	ϕ	I
ϕ	I	ϕ
I	ϕ	ϕ
I	I	I

S_1 \ S_2	ϕ	I
ϕ	I	ϕ
I	ϕ	I

Note that our requirement, "reversing either switch reverses the bulb," fixes the L_1 column (or the inside of the table). However, we do have the alternative of specifying that the upper entry in that column should be ϕ. If we did that, our controls would still work, but according to these displays:

S_1	S_2	L_1
ϕ	ϕ	ϕ
ϕ	I	I
I	ϕ	I
I	I	ϕ

S_1 \ S_2	ϕ	I
ϕ	ϕ	I
I	I	ϕ

In practice, a wiring (or schematic) diagram is made to represent these displays. It will be helpful to compare such a diagram from an electricians' handbook with our displays. The displays can be interpreted for valved pipelines for liquid flow, but the piping configurations will be somewhat different.

You should note that in this early case a peculiar type of symmetry manifests itself. In either of the two pairs of displays above, the labels S_1, S_2, and L_1 may be rearranged in any way while the columns remain fixed, without disturbing the validity and interpretation of the representation. This symmetry will be present only in some representations, not in all.

What you have here are rudimentary parts of a much broader, highly sophisticated, and enormously effective system of analyzing wiring needs and designing wiring configurations to implement controls and operations of

various types. This system has been indispensable to the development of computers of various types, including control systems, and of the great communications networks that now span the earth and reach out into space. The references provide further information and experience.

BIBLIOGRAPHY

Adelfio, Salvatore A., and Christine F. Nolan. *Principles and Applications of Boolean Algebra.* New York: Hayden Book Co., 1964.

Allendoerfer, C. B., and C. O. Oakley. *Principles of Mathematics.* New York: McGraw-Hill Book Co., 1955, 1963.

Andree, Richard V. *Modern Abstract Algebra.* New York: Holt, Rinehart & Winston, 1958.

Birkhoff, Garrett, and Saunders MacLane. *A Survey of Modern Algebra.* New York: Macmillan Co., 1941.

Boole, George. *An Investigation into the Laws of Thought.* New York: Dover Publications, 1951.

———. *The Mathematical Analysis of Logic.* Oxford: Basil Blackwell, 1948.

Bowran, A. P. *A Boolean Algebra.* London: Macmillan & Co., 1965.

Carvallo, Michel. *Monographie des treillis et algèbre de Boole.* Paris: Gauthier-Villars & Co., 1962.

———. *Principes et applications de l'analyse Booléenne.* Paris: Gauthier-Villars & Co., 1965.

Casanova, Gaston. *L'Algèbre de Boole.* Paris: Presses Universitaires de France, 1967.

Eves, Howard, and Carroll V. Newsom. *An Introduction to the Foundations and Fundamental Concepts of Mathematics.* New York: Rinehart & Co., 1958.

Goodstein, R. L. *Boolean Algebra.* New York: Pergamon Press, 1963.

Kaye, Douglas. *Boolean Systems.* London: Longmans, Green & Co., 1968.

Kuntzmann, J. *Fundamental Boolean Algebra.* London: Blackie & Son, 1967.

10

The Imaginary and the Infinite in Geometry

Julius H. Hlavaty

A Euclidean Paradox

A baffling demonstration

Consider a circle O with radius r and a point P *inside* the circle. Find a point P', *outside* the circle, such that

$$OP \cdot OP' = r^2.$$

(Such a point is constructible even with straightedge and compasses only, as seen in figure 10.1.)

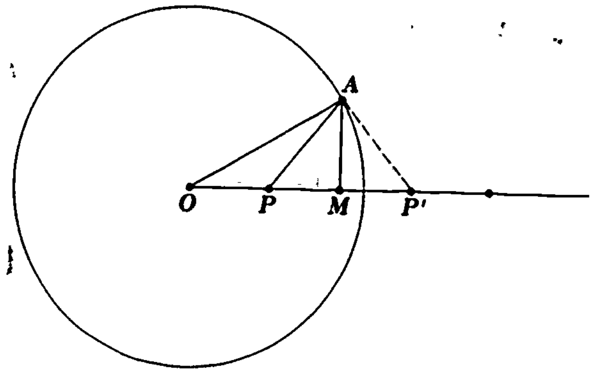

Fig. 10.1

Where do you think P' is located; inside, outside, or on the circle? Let's see!

Find the midpoint M of segment PP'—assume it is inside the circle. Construct a perpendicular to $\overleftrightarrow{OP}$ at M and extend it until it meets the circle in A. Draw $\overline{AO}$, $\overline{AP}$, $\overline{AM}$, and $\overline{AP'}$.

Now, in $\triangle AOM$:

$$OM^2 + AM^2 = AO^2 = r^2 \tag{1}$$

and in $\triangle APM$:

$$PM^2 + AM^2 = AP^2. \tag{2}$$

From equation (1):

$$AM^2 = r^2 - OM^2.$$

Substituting in equation (2):

$$\begin{aligned} PA^2 &= PM^2 + r^2 - OM^2 \\ &= r^2 - (OM^2 - PM^2) \\ &= r^2 - (OM + PM)(OM - PM). \end{aligned}$$

But, since $PM = P'M$:

$$\begin{aligned} PA^2 &= r^2 - OP' \cdot OP \\ &= r^2 - r^2 = 0! \end{aligned}$$

This means that the distance between a point *on* the circle (A) and a point *within* the circle (P) is zero! Does this mean that the two points are coincident?

An explanation?

It is obvious that there is something wrong with the demonstration in the foregoing section. Is it that point M is *outside* the circle and that, therefore, the perpendicular bisector does not intersect the circle?

Let's try coordinates and a special case (the treatment can easily be made general).

Let the circle have its center at the origin and let its radius be 1. Its equation then will be

$$x^2 + y^2 = 1.$$

Let the point P have coordinates $(3/4, 0)$.

Then, to find P' we get $3/4 \cdot a = 1$ and $a = 4/3$, or the point P' has coordinates $(4/3, 0)$.

Using the midpoint formula we find that the coordinates of the point M are $(25/24, 0)$. It is, therefore, outside the circle, and the perpendicular to the x-axis at M does not intersect the circle. Does this solve our paradox?

The line perpendicular to the x-axis at M has the equation $x = 25/24$, and we can look for the intersection of this line with the circle by attempting to solve the system of equations:

$$\begin{aligned} x^2 + y^2 &= 1 \\ x &= \frac{25}{24}. \end{aligned}$$

We get

$$\begin{aligned} \frac{625}{576} + y^2 &= 1 \\ y^2 &= \frac{-49}{576} \\ y &= \frac{\pm 7i}{24}, \end{aligned}$$

so the perpendicular bisector of $\overline{PP'}$ meets the circle in the points (25/24, $7i/24$) and (25/24, $-7i/24$).

Let A be the point (25/24, $7i/24$) and let us find the distance between A and P. Yes, the distance again turns out to be zero!

So, *again* we find that distance AP is zero, that is, that A and P would seem to coincide.

We now have an alternative: We can decide that the algebraic demonstration shows that there simply isn't such a point A on the circle (because we cannot represent points with imaginary coordinates in the plane), or we can adventure forth to see what consequences would follow if we allow "imaginary" points into our work in geometry.

Problems to investigate

1. Can you prove by methods of demonstrative plane geometry that point M must fall outside?
2. Can you make the analytic proof general, that M must fall outside the circle?

Imaginary Points and Lines

Some definitions and their implications

Consider the set of ordered pairs (a, b) where a and b are real or imaginary numbers. We shall call any such pair the coordinates of a point in the plane. We shall even say that any such ordered pair *is* a point in the plane. If a or b or both are imaginary, we shall call the point imaginary.

We shall consider equations with real and imaginary coefficients. We shall assume (though it can be proved) that the Pythagorean theorem (and therefore the distance formula) holds.

Consider the real point A (0, 0) and the imaginary points B $(i, 1)$ and C $(i, -1)$. Find the lengths of the sides of $\triangle ABC$. Do you find $AB = 0$, $BC = 2, CA = 0$? We have an isosceles (?!) triangle, but the triangle inequality does not hold!

Consider the two equations

$$x + iy = 0 \quad \text{and} \quad x - iy = 0.$$

We will say these are two imaginary lines. If we seek their point of intersection we find it to be (0, 0). A real point of intersection for two imaginary lines. As a matter of fact it can be shown (can you show it?) that every imaginary line has one and only one real point on it.

Find any two points on the first of the above equations (that is, two ordered pairs that satisfy the equation.) Find the distance between these two points. Do you find the result interesting? Do you think you would get the same result for any imaginary line?

Problems to investigate

1. Find the midpoints of the line segments determined by the following pairs of points:
   a. $(0, 0), (1, i)$ b. $(1, i), (1, -i)$
   c. $(1, 1), (i, i)$ d. $(1, -1), (-1, 1)$
   e. $(i, -i), (-i, i)$ f. $(5, 7i), (9, 3i)$
   g. $(0, 1 + i), (0, 1 - i)$ h. $(a + bi, 0), (a - bi, 0)$
   i. Make up other problems and solve them.
2. Find the distance between the following pairs of points:
   a. $(0, 0), (i, i)$ b. $(1, i), (1, -i)$
   c. $(i, i), (1, -i)$ d. $(-i, -i), (1, -i)$
   e. $(-i, i), (5, 6)$ f. $(0, 1 + i), (0, 1 - i)$
   g. Make up other problems and solve them.
3. Find the points of intersection of the following pairs of lines:
   a. $x + 2y = 7$. $3x + 6y = 5$.
   b. $x + iy = 1$. $x - iy = 1$.
   c. $2x - 3yi = 1$. $2x + 3yi = 1$.
   d. $x = iy$. $x = 5iy$.
   e. $x + (3 + i)y = 1$. $x - (3 - i)y = 1$.

Homogeneous coordinates

While we could continue our investigation of the role of imaginaries in geometry by means of the usual Cartesian coordinates, we introduce a new system of coordinates that will give more depth and generality to our exploration and will also enable us to talk about the infinite in geometry.

If the Cartesian coordinates of a point A are (x, y) let

$$x = \frac{x_1}{x_3} \quad \text{and} \quad y = \frac{x_2}{x_3},$$

that is,

$$x_1 = x \cdot x_3 \quad \text{and} \quad x_2 = y \cdot x_3,$$

where x_3 is an arbitrarily chosen number.

Then (x_1, x_2, x_3) will be called the homogeneous coordinates of the point A. For example, if A has the Cartesian coordinates $(3, 5)$, the homogeneous coordinates could be $(3, 5, 1)$ or $(6, 10, 2)$ or $(3r, 5r, r)$ where r is any number, real or imaginary. It is clear that any pair of coordinates in the Cartesian system can be transformed into homogeneous coordinates—except that we shall exclude the triple $(0, 0, 0)$. (Why?)

It is also clear that any triple of numbers (a_1, a_2, a_3) will give us a point in the plane (again excluding $(0, 0, 0)$). For example: the triple of numbers $(0, 0, 1)$ or $(0, 0, 7)$ could be the coordinates of the point $(0, 0)$ in ordinary Cartesian coordinates; or $(1, 2, 1)$, $(2, 4, 2)$ and $(r, 2r, r)$ could be the homogeneous coordinates of the point $(1, 2)$.

If the ratios $a_1:a_2:a_3$ are complex, we shall say that we have an imaginary point.

What about homogeneous coordinates (x_1, x_2, x_3) if $x_3 = 0$? It is clear that the ratios x_1/x_3 and x_2/x_3 would approach infinity as x_3 approaches zero. We will therefore *define* all such triples of numbers (for which $x_3 = 0$) as coordinates of points at infinity. Since $x_3 = 0$ is an equation that gives us all such "points at infinity" and since $x_3 = 0$ is a linear equation, we shall say that $x_3 = 0$ is the equation of the line at infinity.

With homogeneous coordinates the "origin" of the ordinary Cartesian coordinate system could be $(0, 0, 1)$ or $(0, 0, r)$, where r is any number. However, we now find *three* origins with our new coordinate system. The other two are: $(0, 1, 0)$ and $(1, 0, 0)$—and we can think of these additional "origins" as points of intersection of the usual x- and y-axes with the line at infinity.

In fact we can think of this new coordinate system as consisting of a triangle with vertices $(0, 0, 1)$, $(0, 1, 0)$, $(1, 0, 0)$ and with three sides having the equations: $x_1 = 0$, $x_2 = 0$, and $x_3 = 0$.

Problems to investigate

1. Find homogeneous coordinates for the following points in the Cartesian coordinate system:
   a. $(0, 1)$ *b.* $(1, 0)$
   c. $(1, 1)$ *d.* $(0, i)$
   e. $(i, 0)$ *f.* (i, i)
   g. $(i, -i)$ *h.* $(1 + i, 1 - i)$
   i. $(1, i)$ *j.* $(-1, -i)$
   k. Other points of your choosing.

2. Find the Cartesian coordinates for the points given with the following homogeneous coordinates:
   a. $(1, 1, 1)$ *b.* $(1, 0, 1)$
   c. $(1, 1, 0)$ *d.* $(0, 1, 1)$
   e. $(1, i, 1)$ *f.* $(i, i, 1)$
   g. $(1, -i, 1)$ *h.* $(i, -1, i)$
   i. $(1, i, 0)$ *j.* $(0, 0, i)$
   k. Other points of your own choosing.

Geometry in the Complex Projective Plane

The linear equation

The equation $Ax + By + C = 0$ or, in homogeneous coordinates, $Ax_1 + Bx_2 + Cx_2 = 0$ is a linear equation. It is a theorem in analytic geometry that this equation represents a straight line. We can define a straight line, then, as the set of ordered triples that satisfy the second form of this equation.

Every imaginary line has one and only one real point on it. To see this, consider the equation

$$Ax_1 + Bx_2 + Cx_3 = 0$$

in which

$$A = a_1 + ia_2,$$
$$B = b_1 + ib_2,$$
and
$$C = c_1 + ic_2.$$
Then
$$(a_1 + ia_2)x_1 + (b_1 + ib_2)x_2 + (c_1 + ic_2)x_3 = 0.$$
This equation will be satisfied by real values of x_1, x_2, x_3 if and only if
$$a_1x_1 + b_1x_2 + c_1x_3 = 0.$$
and
$$a_2x_1 + b_2x_2 + c_2x_3 = 0.$$
These equations have the solution
$$x_1 : x_2 : x_3 = b_1c_2 - b_2c_1 : c_1a_2 - c_2a_1 : a_1b_2 = a_2b_1,$$
which gives us the coordinates of a real point (though the point may be at infinity if $a_1b_2 - a_2b_1 \neq 0$).

For example, the imaginary line $5x_1 + ix_2 + 3x_3 = 0$ has the real point $(-3, 0, 5)$.

Problems to investigate

1. Find real points on each of the following lines:
   a. $3x + 2iy = 7$. b. $2ix - y + 5 = 0$.
   c. $x + iy - 1 = 0$. d. $ix - iy - 1 = 0$.
   e. $x + y - i = 0$. f. $(a + bi)x + (a - bi)y = 0$.
2. Find the equation of a real line through each of the following imaginary points:
   a. $(1, i)$ b. $(1, -i)$
   c. $(i, -i)$ d. $(1 + i, 1 - i)$
   e. (i, i) f. $(2 + \sqrt{3}\,i, 2 - \sqrt{3}\,i)$
   g. $(i, 0)$ h. $(5, i)$
3. Find the real point on each of the following lines:
   a. $x_1 + ix_2 = x_3$. b. $ix_1 + x_2 + x_3 = 0$.
   c. $x_1 + x_2 - ix_3 = 0$. d. $x_1 + (1 + i)x_2 + x_3 = 0$.
   e. $(1 + i)x_1 + (1 - i)x_2 + x_3 = 0$. f. $ix_1 - ix_2 + x_3 = 0$.
   g. $x_2 = mx_1 + ibx_3$.

Two parallel lines intersect at infinity. Consider the equations of two lines:
$$A_1x_1 + B_1x_2 + C_1x_3 = 0 \quad \text{and} \quad A_2x_1 + B_2x_2 + C_2x_3 = 0;$$
these lines are real or imaginary, depending on the values of the coefficients. They have the unique solution:
$$x_1 : x_2 : x_3 = B_1C_2 - B_2C_1 : C_1A_2 - C_2A_1 : A_1B_2 - A_2B_1.$$

This solution represents the point common to the two lines. If $A_1B_2 = B_1A_2$, the point of intersection is on the line at infinity—$x_3 = 0$—and the lines are parallel. If the lines are imaginary, we shall call them parallel by definition. We have, therefore, proved: Two straight lines intersect in one and only one point. If the lines are parallel, the point of intersection is on the line at infinity.

For example, the equations:

$$x_2 = x_1 + 3x_3 \quad \text{and} \quad x_2 = x_1 - 5x_3$$

have the solution (1, 1, 0). These lines are parallel and intersect the line at infinity, $x_3 = 0$, in the point (1, 1, 0).

Problem to investigate

1. Find the point of intersection of each of the following pairs of lines:
   a. $x_1 = i, x = -i$
   b. $x_1 = ix_3; x_1 = i - ix_3$
   c. $x_2 = ix_3, x_2 = -ix_3$
   d. $(1 + i)x_1 = (1 - i)x_3, (1 - i)x_1 = (1 + i)x_3$
   e. $x_1 + x_2 + x_3 = 0, x_1 + ix_2 + ix_3 = 0$

Minimal (or isotropic) lines. Let us consider the lines that connect the origin of the Cartesian coordinate system, (0, 0, 1), with the imaginary points $(1, i, 0)$ and $(1, -i, 0)$.

The equations of these lines are $x_2 = ix_1$ and $x_2 = -ix_1$, and their slopes are i and $-i$. Since the distance between any two points on either line is zero (see the section above entitled "Some definitions and their implications"), these lines are called *minimal* or *isotropic* lines.

If we raise the general question of what is the locus of points at zero distance from a given point (x_0, y_0) we find, using ordinary Cartesian coordinates:

$$(x - x_0)^2 + (x - y_0)^2 = 0$$

or

$$((x - x_0) + i(y - y_0))\cdot((x - x_0) - i(y - y_0)) = 0.$$

so

$$(x - x_0) + i(y - y_0) = 0$$

or

$$(x - x_0) - i(y - y_0) = 0.$$

In homogeneous coordinates these equations are of the form:

$$x_1 + ix_2 + kx_3 = 0 \quad \text{or} \quad x_1 - ix_2 + mx_3 = 0.$$

These are, in fact, the minimal lines that go through the given point. (Another way of thinking of the minimal lines is that they are lines with slope i or $-i$.) In other words, the locus of points at zero distance from a given point consists of the given point and the two minimal lines going through that point.

The circle

Adapting ordinary Cartesian coordinate equations to homogeneous coordinates, we can consider the equation

$$x_1^2 + x_2^2 = 0$$

as the equation of a circle with center at the origin and radius 0.

If one or more coefficients are complex, the line will be said to be imaginary. The equation $x_3 = 0$ will be called the line at infinity.

We see immediately that this equation reduces to the equations:

$$x_2 = ix_1 \quad \text{and} \quad x_2 = -ix_1,$$

in other words, to the two minimal lines through the origin. We can say, therefore, that the null circle consists of the center and of the two minimal lines going through the center.

It is clear that the null circle given above intersects the line at infinity, $x_3 = 0$, in the points $(1, i, 0)$ and $(1, -i, 0)$. (Check this statement.)

We shall now show that every circle in the plane goes through these two points at infinity. For this reason the points $(1, i, 0)$ and $(1, -i, 0)$ are called the *circle points at infinity.*

The equation of an arbitrary circle, in homogeneous coordinates (recall and transform the ordinary Cartesian equation), is

$$x_1^2 + x_2^2 + a_1x_1x_3 + a_2x_2x_3 + a_3x_3^2 = 0.$$

A direct substitution will show that the points $(1, i, 0)$ and $(1, -i, 0)$ satisfy this equation.

Moreover, we can show that any conic section that goes through the two circle points at infinity is a circle. The general equation of a conic, in homogeneous coordinates, is

$$Ax_1^2 + Bx_1x_2 + Cx_2^2 + Dx_1x_3 + Ex_2x_3 + Fx_3^2 = 0.$$

If this equation is satisfied by $(1, i, 0)$ and $(1, -i, 0)$, then

$$A + Bi - C = 0 \quad \text{and} \quad A - Bi - C = 0.$$

Solving these equations, we see that $A = C$ and $B = O$, and we get

$$Ax_1^2 + Ax_2^2 + Dx_1x_3 + Ex_2x_3 + Fx_3^2 = 0,$$

which is the equation of a circle.

The other conic sections

From ordinary geometry we learn that the intersections of a plane and a conic surface yield the conic sections—circle, ellipse, hyperbola, and parabola. (Of course, there are special, degenerate cases in which these sections are points, two intersecting lines, two parallel lines, or two coincident lines.)

What light is shed on the conics by the techniques we have introduced—homogeneous coordinates, imaginary points and lines, and points and lines at infinity?

The parabola. In analytic geometry we learn that the simplest form of the equation of a parabola is

$$y = cx^2 \quad \text{or} \quad x_2x_3 = cx_1^2.$$

The parabola intersects the line at infinity, $x_3 = 0$ in the point $(0, 1, 0)$—in fact it intersects that line in two coincident points $(0, 1, 0)$ and $(0, 1, 0)$—and therefore we can say that the parabola is tangent to the line at infinity at that point. The point of tangency is the point in which the axis of symmetry of the parabola intersects the line at infinity.

The ellipse. We learn in analytic geometry that the equation:

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$$

represents an ellipse. Let us extend this definition by considering each of the following as an ellipse, also:

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = -1 \quad \text{and} \quad \frac{x^2}{a^2} + \frac{y^2}{b^2} = 0.$$

Let us change these equations to homogeneous coordinates:

$$\frac{x_1^2}{a^2} + \frac{x_2^2}{b^2} = x_3^2, \quad \frac{x_1^2}{a^2} + \frac{x_2^2}{b^2} = -x_3^2, \quad \frac{x_1^2}{a^2} + \frac{x_2^2}{b^2} = 0.$$

Clearly, the last equation becomes the two conjugate, imaginary, intersecting lines:

$$x_1b - iax_2 = 0 \quad \text{and} \quad x_1b + iax_2 = 0.$$

The lines intersect in the point $(0, 0, 1)$ (the origin of the Cartesian coordinate system), and they intersect the line at infinity in the conjugate-imaginary points: $(a, bi, 0)$ and $(a, -bi, 0)$.

We say, therefore, that the third equation represents degenerate ellipse that consists of two conjugate-imaginary lines that intersect in a real point, the origin, and intersect the line at infinity in two conjugate-imaginary points.

The second equation $(x^2_1/a^2 + x^2_2/b^2 = -x^2_3)$ is an ellipse with no real graph, but it also intersects the line at infinity in the points $(a, bi, 0)$ and $(a, -bi, 0)$.

The equation $x^2_1/a^2 + x^2_2/b^2 = x^2_3$, which represents a real ellipse, intersects the line at infinity in the points $(a, bi, 0)$ and $(a, -bi, 0)$.

In the special case where $a = b = 1$ the equations reduce to

$$x_1^2 + x_2^2 = x_3^2, \quad x_1^2 + x_2^2 = x_3^2, \quad \text{and} \quad x_1^2 + x_2^2 = 0,$$

the equations of a real circle, an imaginary circle, and the point-circle, respectively. Each of these intersects the line at infinity in the points $(1, i, 0)$ and $(1, -i, 0)$—the circle points at infinity. This confirms what we observed in an earlier section on the circle.

Also, in the last case, the circle reduces to the conjugate isotropic lines $x_1 + ix = 0$ and $x_1 - ix_2 = 0$ and their real point of intersection $(0, 0, 1)$.

The hyperbola. The equations

$$\frac{x_1^2}{a^2} - \frac{x_2^2}{b^2} = x_3^2 \quad \text{and} \quad \frac{x_1^2}{a^2} - \frac{x_2^2}{b^2} = 0$$

represent, respectively, real nondegenerate hyperbolas and real degenerate hyperbolas (if a and b are real).

The second of these reduces to the two real intersecting lines:

$$bx_1 - ax_2 = 0 \quad \text{and} \quad bx_1 + ax_2 = 0.$$

These lines intersect the line at infinity in the two real points $(a, b, 0)$ and $(a, -b, 0)$, and they intersect each other at the point $(0, 0, 1)$ (the origin of the Cartesian coordinate system).

The real, nondegenerate hyperbola,

$$\frac{x_1^2}{a^2} - \frac{x_2^2}{b^2} = x_3^2,$$

also intersects the line at infinity in the two real points $(a, b, 0)$ and $(a, -b, 0)$.

BIBLIOGRAPHY

Adler, Claire Fisher. *Modern Geometry: An Integrated First Course*, pp. 110-16. New York: McGraw-Hill Book Co., 1958.

Adler, Irving. *A New Look at Geometry*, pp. 305-11. New York: John Day Co., 1966.

Borel, Emile, and Robert Deltheil. *La Géométrie et les imaginaires*. Paris: Editions Albin Michel, 1950.

———. *L'Imaginaire et le réel en mathématique et en physique*. Paris: Editions Albin Michel. 1952.

Graustein, William Casper. *Introduction to Higher Geometry* pp. 29-34, 115-36. New York: Macmillan Co., 1930.

Maxwell, Edwin Arthur. *Methods of Plane Projective Geometry Based on the Use of General Homogeneous Coordinates*, pp. 1-6. New York: Cambridge University Press, 1946.

Meserve, Bruce E. *Fundamental Concepts of Geometry*, pp. 18-20, 95-100, 27-37. Reading, Mass.: Addison-Wesley Publishing Co., 1955.

Woods, Frederick S. *Higher Geometry*, pp. 2-4, 27-33. Boston: Ginn & Co., 1922.

Yaglom, I. M. *Complex Numbers in Geometry*. New York: Academic Press, 1968.